


```
OY 181 RRYSEYGYTYITITKLDKNGIQIWLPOFKSDARVDLMLRPTGGGTYIGRNSVDMCFY 240
| | | | |
DB 181 RRYSEYGYTYITITKLDKNGIQIWLPOFKSDARVDLMLRPTGGGTYIGRNSVDMCFY 240
OY 241 DGSTNSSSLEIFRQDNNPKSDGKFLRKINDOTKEIATYLSLLAGKSLTPNGTSLNT 300
| | | | |
DB 241 DGSTNSSSLEIFRQDNNPKSDGKFLRKINDOTKEIATYLSLLAGKSLTPNGTSLNT 300
OY 301 ADAASLETNNRITATVATMPETISVPVLCMPGRLOLDKAVENPEAGOVYNGINVTFTPSQOT 360
| | | | |
DB 301 ADAASLETNNRITATVATMPETISVPVLCMPGRLOLDKAVENPEAGOVYNGINVTFTPSQOT 360
OY 361 L 361
|
DB 361 L 361

RESULT 2
O47115 PRELIMINARY; PRT; 363 AA.
ID 047115; 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Cooc precursor.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94344028; PubMed=7915003;
RA Froehlich B., Karakashian A., Melisen L.R., Wakefield J.C., Scott J.R.;
RT "Cooc and Cood are required for assembly of CSI pili.";
RL MOL. Microbiol. 12:387-401(1994).
EMBL: X76908; CAA54230.1; -
DR InterPro: IPR001412; TRNA-synL.
DR PROSITE: PS00178; AA_TRNA_LIGASE_1; UNKNOWN_1.
KW Signal.
FT SIGNAL. 1 18 POTENTIAL.
FT CHAIN 19 363 COOD.
SQ SEQUENCE 363 AA; 40139 MW; AA948BAADDED5F72 CRC64;

Query Match 50.3%; Score 949.5; DB 2; Length 363;
Best Local Similarity 53.1%; Pred. No. 5.8e-65;
Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;

3 KILFIETLFSSVLTFAVSADKIPGDE--SITNIF-GPR-DRNESSPKHNLNNHITAY 58
| | | | |
2 KIIFFIFL---SIFSAVVSAGRPETTVGNLTKFSQAPRLDRSVQSPYINIFNHVAGY 57
OY 59 SESHTLYDRMTFLCLSHNTLNGACPTSENPSSSVS-GETNTLTOTTEKRSILKRELQI 117
| | | | |
DB 58 SLISLSDRLVLTCTSSNPNVNGACPTI---GTSVOYGTFTTTLQTEKRSILKRNINL 114
OY 118 KGKYLLEFKSVNCPG--LTLNSAHFNCNKNA-ASGASLYLIPAGELKMLPFGINDAT 174
| | | | |
DB 115 AGKKKPIWENQODFSMLWNLNSKMSGCAHGANGLMLWYIPAGINKLPRGIMEAT 174
OY 175 LKLRVRRYSET---YGYTITNTIKLTDKNGIQIWLPOFKSDARVDLMLRPTGGGTYI 230
| | | | |
DB 175 LILRLS-RYGEVSTHGNVTYITVDLTDKNGIQIWLPGFHSNPRVDLMLRPGNKYS 233
OY 231 GRNSVMCFEYDGYSTNSSSLEIFRQDNNPKSDGKFLRKINDOTKEIATYLSLLAGKSL 290
| | | | |
DB 234 GSNLSLMCFYDGYSTNSDVAIKFQDNNPTNSSEYNLYKYG-GTEKLPYAVSLMGKIF 292
OY 291 TPNNGTSLNTADASLETNNRITATVATMPETISVPVLCMPGRLOLDKAVENPEAGOVYNGI 350
| | | | |
DB 293 YPNNGSFTINDSSVLETNMNRITATVAMPEVNVYVLCMPARLLMLADYNAAPDAQYSGOI 352
OY 351 NVTFTPSQOTL 361
```

```
DB 353 YITFTPSVENTL 363
| | | | |

RESULT 3
O47119 PRELIMINARY; PRT; 364 AA.
ID 047119; 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE Cood precursor.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C91F-6;
RX MEDLINE=96071908; PubMed=7591145;
RA Froehlich B.J., Karakashian A., Sakellaris H., Scott J.R.;
RT "Genes for CS2 pili of enterotoxigenic Escherichia coli and their
interchangeability with those for CSI pili.";
RL Infect. Immun. 63:4849-4856(1995).
EMBL: Z47800; CAA87763.1; -
KW Signal.
FT SIGNAL. 1 18 POTENTIAL.
FT CHAIN 19 364 COOD.
SQ SEQUENCE 364 AA; 40979 MW; ECTAC738ADE195B6 CRC64;

Query Match 49.3%; Score 930.5; DB 2; Length 364;
Best Local Similarity 48.4%; Pred. No. 1.7e-63;
Matches 177; Conservative 64; Mismatches 118; Indels 7; Gaps 3;

OY 1 MNKILFIETLFSSVLTFAVSADKIPGDESTINIGPRDRN--ESSPKHNLNNHITAY 58
| | | | |
DB 1 MKRVIFVLSMFLCSOYVGOQSMHTNVEAGSINKTESIGPIDRSAASYPAHYIFHEHVAGY 60
OY 59 SESHTLYDRMTFLCLSHNTLNGACPTSENPSSSVSGETNTLTOTTEKRSILKRELQI 118
| | | | |
DB 61 MKDHSLEFDRMTFLCWSSTDSKACAPTEGNSKSS--QGETNLTITTEKRSILKRLINL 118
OY 119 GYKOLLEFKSVNCPG--PSGLTLNSAHFNCNKNAASGASLYLIPAGELKMLPFGINDATL 175
| | | | |
DB 119 GYKRLYESDRCHYVDKMLNSHTYKCVGSPTRGVDFTLYIPQGEIDGLTGIMEATL 178
OY 176 KLRVRRYSETGYTYITNTIKLTDKNGIQIWLPOFKSDARVDLMLRPTGGGTYIGRNSV 235
| | | | |
DB 179 ELRVKRRHYDNGHTYKVNITVDLTDKNGIQIWLTPKFSHPRIDLMLRPEGNGKYSNVL 238
OY 236 DMCFYDGYSTNSSSLEIFRQDNNPKSDGKFLRKINDOTKEIATYLSLLAGKSLTPNG 295
| | | | |
DB 239 EMCLYDGYSTHOSITTEKFDSDQGTGNNETNLTKTGEPLKLYKLSLLGKEFPNNG 298
OY 296 TSLNIDASLETNNRITATVATMPETISVPVLCMPGRLOLDKAVENPEAGOVYNGINVTFT 355
| | | | |
DB 299 EAPFTINDSSSLFINMNRKIVSLPDISIPVLCMPANLTFESELNPEAGYSGILNVTFT 358
OY 356 PSSQOTL 361
|
DB 359 PSSSSL 364

RESULT 4
O9XDS1 PRELIMINARY; PRT; 359 AA.
ID O9XDS1; 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE TSAD protein (putative fimbrial protein).
GN TSAD OR TCED OR STY0348.
```


RT "Isolation of a new pmp sequence and evidence of pmp polymorphism in
RT serotype-1 Chlamydia psittaci strains."
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF243415; AAL36959.1; -
DR InterPro: IPR003368; Chlamydia_PMP.
DR InterPro: IPR003357; OMP.
DR Pfam: PF02415; DUF145; 1.
DR Pfam: PF02385; OMP; 1.
FT NON_TER 700
SQ SEQUENCE 700 AA; 74200 MW; 2C38E504FA8D1D1 CRC64;

Query Match 6.1%; Score 115.5; DB 2; Length 700;
Best Local Similarity 18.3%; Pred. No. 1.3; Indels 139; Gaps 18;
Matches 73; Conservative 63; Mismatches 124;

QY 42 NESSPKHNIINHTA-----YSESHLYDRMTFLCLT---SHMTLNGAC----- 83
DB 27 NEILTSDSYNGVNTDSEFEVKETTSAGIAYTCGNCVCISAGKDSPLNKSCFSETENLS 86
QY 84 -----PISENDSSSVSGETNITLQFTKRSLLKRELOIKGYKOLLFRSVN 129
DB 87 FIGNGYTLCEFDNITTTASNAGAIWVSGD-----QKTLNVSGFS--LFSCAH 130
QY 130 CPSSGLT-----LNSAHF--NCKN-----NAASGASLYIYIP-----A 159
DB 131 CPPTGTGYGAIQTKGVSTFGNNKLIIPDNCSGEGAIKCATGSMELKEGNSYVFS 190
QY 160 GELKNIPLFGIMDATLKLARKRYSEYGYTITIKILDKGMIQIWLPOFKSDARVDL 219
DB 191 GNSGQKGGAIY--TKKLITTTADGPTLFSNNSVAS---SPKGG-AICLDLDTSECLTA 244
QY 220 NL-----RPTGGGTYIGRNSVDM-----CFYDGYSTNSSSLE 251
DB 245 NLGDTFEDGNKVIKTNCGSSIVKRNALIDLGSGKFTKLNAKEGPIFHPDIANTGGSTE 304
QY 252 IRRQDNPKSD-----GR--FYLRKINDTKEIAYTL-----SLLLA 286
DB 305 IEL--NRTESDPTTYTGKIVSEGEKLSDDEKTVDPNLKSYRKQPLKIGAGSLVKDGYTLE 362
QY 287 GKSLPTNGTSLNIADAASLETNMNRITAAVMPESVAV 325
DB 363 AKKITQTKGSTVWDLGTTLOTPTSSSGETTITLMDINI 401

RESULT 7
P71132 PRELIMINARY; PRT; 847 AA.
AC P71132; 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
OS Chlamydia abortus.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.1.
OX NCBI_TaxID=83555;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=526/3;
RX MEDLINE=96406378; PubMed=8810511;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RT "Identification of a multigene family coding for the 90 kDa proteins
of the ovine abortion subtype of Chlamydia psittaci."
RL FEMS Microbiol. Lett. 142:277-281(1996).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=526/3;
RX MEDLINE=98187897; PubMed=9529048;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the
highly immunogenic cluster of 90-kilodalton envelope proteins from the
RT Chlamydia psittaci subtype that causes abortion in sheep."
RL Infect. Immun. 66:1317-1324(1998).
DR EMBL: U65942; AAC15921.1; -.

DR InterPro: IPR003368; Chlamydia_PMP.
DR InterPro: IPR003357; OMP.
DR Pfam: PF02415; DUF145; 1.
DR Pfam: PF02385; OMP; 1.
SQ SEQUENCE 847 AA; 90694 MW; 754C958E7F1179E CRC64;

Query Match 6.1%; Score 115.5; DB 2; Length 847;
Best Local Similarity 21.6%; Pred. No. 1.7;
Matches 94; Conservative 58; Mismatches 153; Indels 131; Gaps 21;

QY 10 LFFSSVFTFAVSADKIPGDESTINIFGPRDRNESSPKHNIINHTAYSESHLYDRMT 69
DB 9 LISSSLIVSNLSLSE--PDQKTLTSAHSYNGNTNSEP-----FNPSTSNSTGYTCG 62
QY 70 FLCLT-----SHNTLNGACP-----SENPSSSVSGETNITL 102
DB 63 NICIATVAGLDGSGSSCFDITAGNLSFLGNGYTLCEFDNITTTGSHGALSIVSG--TNKT-- 120
QY 103 QFTKRSLLKRELOIKGYKOLLFRKSVNCPSSGLTNSA-----HFNCK 145
DB 121 -----LDISGFS--LFSCAICPPGATGYGAIKAVGNTTIKDNLSLVEHKNCSF 166
QY 146 NAASGASLYIYIPAGELX-----NLPR-----GIMDATLKLARKRYSEYGYTIT 192
DB 167 GEGGAIQCKKASSSEAEKTIENNONLVFAENSSSSGGAIYAD--KLITVSGPTLFSNNSV 225
QY 193 NIT-----IKLNDKGMIOIWLPOFKSDARVDLN--LRPTGGGTYIGRNSVDM----- 237
DB 226 SASSPKGAICINDSG--ECSLTADLGDITFDGNKIITKNGSPYTRNSIDLGSGKFT 284
QY 238 -----CFYDGYSTNSSSLEIRFQDNPKSDCK--FYLRKINDTKEIAYTL--- 281
DB 285 KLNAREGFGIFFYDPTITGGSDDELINKQDPTVDT--GKIYSEGERLSDDEKKAVALNKSD 343
QY 282 -----SLLLAG-----KSLPTNGTSLNIADAASLET--NMNRITAAVMPESIV 323
DB 344 FKQPLKIGSGLILKQGVLTLETFSFTQEGATVYMDGTTLOTPTSSSGETTITLMDIN 403
QY 324 PVLCPMPGRLQDAKVE 339
DB 404 ASLGGGVADPAKVE 419

RESULT 8
O807H7 PRELIMINARY; PRT; 865 AA.
AC O807H7; 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=99CMA304;
RX MEDLINE=21849375; PubMed=11860674;
RA Yamaguchi J., Vallari A.S., Swanson P., Bodelle P., Kaptue L.,
RA Nansop C., Zekeng L., Gurtler L.G., Devare S.G., Brennan C.A.;
RT "Evaluation of HIV Type 1 Group O Isolates: Identification of Five
RT Phylogenetic Clusters."
RL AIDS Res. Hum. Retroviruses 18:269-282(2002).
DR EMBL: AF383244; AAL98866.1; -.
FT NON_TER 1
SQ SEQUENCE 865 AA; 97118 MW; 996B83862AA51ACA CRC64;

Query Match 6.0%; Score 114; DB 15; Length 865;
Best Local Similarity 21.5%; Pred. No. 2.2; Indels 134; Gaps 22;
Matches 93; Conservative 56; Mismatches 150;

QY 21 VSADKIPGDESTINIFGPRDRNESSPKHNIINHTAYSESHLYDRMTFL 71

```

Db 75 VFDPDPPEYPLHN-----TDKFNIMWENYAVKQMGDIIDMEQSLKPCVAMTEL 125
OY 72 CUSNHT-INGACPTSENPSSSVS-----GETNITITQFTEKSLIRELOIKGYKOLLK 126
Db 126 CVOAMNCTSLN---DTKSNPTSPENLAKKEFNVTIVVADKKEKKOALPEYSDLMKLDK 182
OY 127 SYVCPSEGLTNSAHFNCKNNNAASGA-----SLXYLPAG-----ELKNLPFG- 169
Db 183 NTNTNTMTLT-----NCNSTTISQACPKVSFEPIPIHYCAPAGYALFKCNINIEFGTGC 238
OY 170 -----IMDATL---KLRY-KRYSSEYV-YYTINITIKLTDK--G 202
Db 239 KNTVVTCTHGAKPTVSTQLINGTLSTKGRIRMTKNIDNVNIYTLNLTSLKTKCRKG 298
OY 203 NQIWLPOKSDARDLNLPTGGGTYIGRNSYDMCFYGGSTNSSLEIRPDNNPKSD 262
Db 299 NMTVOQVSTIGPAMYSMDLR-----GTERMTSRVACEY-----NSTDWE-----RTLKQT 344
OY 263 GKPYLRKINDDKELAVTSLLAG-----KSL-----TPITGSL 298
Db 345 AEFLELVN-NTRKVDMTFSNSSGGDEPVANLHFNCHGEFFYCNTSSLNNTFSNGSTC 403
OY 299 NTADAASLETNNRI-----TAVTMEISVPLYCWPGRQLDPAKVENPPA 343
Db 404 NVTKSN-N-STNNRILPCRLROYVKSWIOGSGLYAPPIRGNLTKMSNITGLMIQMDP--- 460
OY 344 GQYMGNINVTFRP 356
Db 461 ---WNHSMNATFRP 470

RESULT 9
O8XNM2 PRELIMINARY; PRT; 743 AA.
ID O8XNM2
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Hypothetical protein CPE0220.
GN CPE0220.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
STRAIN=13 / TYPE A;
PubMed=11792842;
RA Shima T., Ohlani K., Hirakawa H., Ohshima K., Yamashita A.,
Shima T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
"Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater."
RT Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RL EMBL: AP003185; BAB79926.1;
DR InterPro: IPR001899; Gram_pos_anchor.
DR TIGRfams: TIGR01167; LPTG_anchor.1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 743 AA; 81149 MW; A7C451E9C5A5E545 CRC64;

Query Match 5.8%; Score 111.5; DB 16; Length 743;
Best Local Similarity 18.8%; Pred. No. 2.9; Mismatches 115; Indels 95; Gaps 13;
Matches 64; Conservative 66;

OY 7 IFFLFSSVLETFVAVSADRIPEDESITNIFGRDRNES-----SPKHNILNNHTTYSSES 61
Db 88 LMTVYVNSLSLGGVNMNIEVSAGGEALKEENKDKSITFEVSPDVKVYIGLFTIMGKR 147
OY 62 HTLYDRMTFLCSSHNTLN--GACPTSENPSSSV-----SG-----ETNITIQ 103
Db 148 VELF-----LVNDMTVWILDEAPITNNAKDISVYOGDAIDLLSGVIGTDKEDSNLKYE 201

```

```

OY 104 FTEKRSLIK-RELOIKGYKQLFKSVNCPSEGLTNSAHFNCKNNNAASGASLYIYPAGEL 162
Db 202 ISGDTSFVKDGAELPQVPIITYK-----YDSSGQFD----- 234
OY 163 KNLPEGIDATLKLKRVKRRYSSEYTYTINITIKLTGKNIO-----IWLPOKSDARD 218
Db 235 -----EKVYVNTVNNKKTTLGDSGYTLKNTVQYVVOGGMNETGSMARKYLSDSRID 285
OY 219 LNIPTGGGTYIGRNSYDMCFYDGYSTNSSLEIRPDNNPKSDGKPYLRKINDDKELA 278
Db 286 IS-----NCKNTVTLTF-----NSELYAFLEKNNVTVDEKEVAEYVKDKRITK 329
OY 279 YTL-----SLLAGSLRPTNGTSLINDAASLEIN 309
Db 330 FNIPDLSIDIVSTLVSMGKEYSEKTYTLNIDYAKKLEDN 369

RESULT 10
P71133 PRELIMINARY; PRT; 846 AA.
ID P71133
AC P71133;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE POMP91B precursor.
OS Chlamydia abortus.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83555;
RN [1]
RP SEQUENCE FROM N.A.
STRAIN=526/3;
RC MEDLINE=96406378; PubMed=8810511;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RT "Identification of a multigene family coding for the 90 kDa proteins
of the ovine abortion syndrome of Chlamydia psittaci."
RL FEMS Microbiol. Lett. 142:277-281(1996).
RN [2]
RP SEQUENCE FROM N.A.
STRAIN=526/3;
RC MEDLINE=98187897; PubMed=9529048;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the
highly immunogenic cluster of 90-kilodalton envelope proteins from the
Chlamydia psittaci subtype that causes abortion in sheep."
RL Infect. Immun. 66:1317-1324(1998).
RL EMBL: U65943; AAC15923.1;
DR InterPro: IPR003368; Chlamydia_PMP.
DR InterPro: IPR003357; OMP.
DR Pfam: PF02415; DUF145; 1.
DR Pfam: PF02385; OMP; 1.
KW Signal.
FT CHAIN 1 16 POTENTIAL.
FT SIGNAL 17 846 POMP91B.
SQ SEQUENCE 846 AA; 90834 MW; 4CDC31DC0C32964E CRC64;

Query Match 5.8%; Score 110; DB 2; Length 846;
Best Local Similarity 21.9%; Pred. No. 4.4;
Matches 86; Conservative 51; Mismatches 132; Indels 124; Gaps 20;

OY 42 NESSPKHNILNNHTA-----YSESHLYDMMTFCLTS-----SHNTLNAC--PSENS 90
Db 27 NETLSSDSYNGNVTSDPEFVAKETTSYGALITCEGNVCISYAGKDSFLNKSSEETENLS 86
OY 91 SSVSGET-----NITLQFTEKSL-----IKRELOIKGYKQLFKSVNCPSEGLTNSA--- 139
Db 87 FLG-NGYTLCEFNITQSSHGPAISVSGTNKTLDISGF-SLPSCAVCCPPTGTGAIOT 144
OY 140 -----FNCN-----KNAASGASLYIYPAGELKLPF----- 167
Db 145 KQTTLLKDNSSLVFNKNCSTAGAGAQCKSSSTAELKL-----ENNNKLVFSENSKEKG 200
OY 168 GGIMDATLKL-----RVKRRYSSEYTYTINITIKLTGKNIOIWLPOKSDAR 216

```

Db 201 GAIVADLTIVSGPPLFSSNNSVSHNSPKGCA-----ICIKDSGC-----ECSLTANLGDIT 253
QY 217 VDLN-LRPTGGGTYIGRNSVDY-----CFYDGYSTNSSLEIRFODN 257
Db 254 FDGKITTNGSGPPTVRNSIDGSGGKPKLNKKEGSGIFFFYPIANTGSTIEL--N 311
QY 258 NPKSD---GK--FYLRKINDKREIAYTL-----SLLAGKSLTP 292
Db 312 KTESDYYTKRYFSGEKLSEDEKTPANLKSYPKOPKIGAGSLVLKDVLTLEAKKITQ 371
QY 293 TNGSLNIAAASLETWNMRITAVTMEISVYP 325
Db 372 TKGSTVMDLGTTLQTPSSSGETITLNLINDI 404
RESULT 11
Q8T2G3 PRELIMINARY; PRT; 2454 AA.
AC 08T2G3
ID 01-JUN-2002 (Tremblrel. 21, Created)
DE 01-JUN-2002 (Tremblrel. 21, Last sequence update)
OS 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 274.1 kDa protein.
OC Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX4:
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tungann B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
HL "Sequence and Analysis of Chromosome 2 of Dictyostelium."
DB Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
KW EMBL: AC115608; AAL92375.1; -
KW Hypothetical protein 2454 AA; 274123 MW; B594001B4487C724 CRC64;
SQ SEQUENCE 2454 AA; 274123 MW; B594001B4487C724 CRC64;
Query Match 5.8%; Score 109.5; DB 5; Length 2454;
Best Local Similarity 23.0%; Pred. No. 19;
Matches 79; Conservative 42; Mismatches 141; Indels 81; Gaps 18;
QY 42 NESSPKNIILNHTTAVSESHLYDRMFLCLSSHTNLGACPTSENPSSSV-SGETNI 100
Db 996 NNSGSSGNLNNMTPTSSQTN-----SSSTTAGTNPSTSTSTGTGTNSTTT 1045
QY 101 TLOPTEKRSIKRELQKGYKQLLFKSVNCPG-----LTLS--AHFNCNNAASGA 151
Db 1046 TTNEPEIPDLKK-----MEFLLPQNGVILPFLQSNFTHEHVKQMLKFL 1094
QY 152 S-LYLYIPAGELKNLPFGIWDATLKLVRKRYSEYGYTINITIKL-----TDKGNIQ 205
Db 1095 NHVFLF-----TKKKLYIDSYISYIFIKLYNNYIPDQATLT 1138
QY 206 IWLPOFKSDARVDLNRPTGGGTYIGRN-SVDMCFYD-----GISTNSSLEIRFODN--- 257
Db 1139 L-----ORLFLNLIISFS--VNKNDKISMKFYOLRTMDFLVRVNELEYEIKQMKDQ 1187
QY 258 --NPKSDGKFLRKINDKREIAYTL--SLLAGKSLTPNTNGSLNIAAASLETWNMTA 315
Db 1188 FLKPLSSIPFFANS-SDDTSSSSSTSSSL-----SLSTTTTTTTTTTTKTTTNTNSTTP 1242
QY 316 V-TMPEISVPLCWPG-RLOLDKAVENPEAGQYMGINIVTTP 356
Db 1243 IKTGGPPTPKLAPGTGKLNGLNSGITRTDSSNNNNNTSSP 1285
RESULT 12
Q8T7V5 PRELIMINARY; PRT; 807 AA.
AC 08T7V5
ID 01-JUN-2002 (Tremblrel. 21, Created)
DE 01-JUN-2002 (Tremblrel. 21, Last sequence update)
OS 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Nicotinic acetylcholine receptor Dalpha5 subunit.
GN NACRALPHA-34E.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21969411; PubMed=11973307;
RA Grauso M., Reenan R.A., Culetto E., Sattelle D.B.;
RT "Novel Putative Nicotinic Acetylcholine Receptor Subunit Genes,
RT Dalpha5, Dalpha6 and Dalpha7, in Drosophila melanogaster Identify a
RT New and Highly Conserved Target of Adenosine Deaminase Acting on RNA-
RT Mediated A-to-I Pre-mRNA Editing."
RL Genetics 160:1519-1533 (2002).
DR EMBL: AF272778; AAM13390.1; -
KW Receptor.
KW SEQUENCE 807 AA; 91223 MW; C8B4F6B34287C8C8 CRC64;
SQ SEQUENCE 807 AA; 91223 MW; C8B4F6B34287C8C8 CRC64;
Query Match 5.8%; Score 109; DB 5; Length 807;
Best Local Similarity 20.2%; Pred. No. 5;
Matches 78; Conservative 48; Mismatches 140; Indels 120; Gaps 18;
QY 9 TLFFSVLTFEAVSADKIPDESITNIFGPRDRNESSPKHNIILNH----- 54
Db 249 TIAFISYLSGFAQLKNSSSSSSSN-----SSNNSQIILNGKHSWIFLLIYNLSAK 304
QY 55 --ITAVSESHLYDRMFLCLSSHTNLGACPTSENPSSSVSGETNITLQTEKNSLK 112
Db 305 VCLAGYHEKRLH-----LLDPNTLERVNLNESPQLQSLFGLTLMQIIDVDEKQLLV 359
QY 113 RELQIKGYKQLLFKSVNCPGSLTLNSAHFNCNNAASGASLYLYIPAGELKNLPFGIWD 172
Db 360 TNWMLK-----LEWDMNLRKMTSDYGVV-----DLR-IPPRIRWK 395
QY 173 ATLKLVRKRYSEF---YGYTINITIKLTDKGNIDWLPQ---FKSDARVDLNRPTGG 226
Db 396 PDVLM-----YNSADEGFGDTGYQNVVR--NNGSC-LVYPPGKFKCTCKIDLTWPF--- 444
QY 227 GTYIGRNSVDMCF---YDGYSTNSSLEIRFODNPKS-----DGKFLRKINDDTKE 276
Db 445 --FDDQRCMKRGKSWYDGF---QLDLQDDEGGDISVYLVNGEWELLPVGRKNE 496
QY 277 IAYT-----LSLLAGKSLTPNTNGSLNIAA 303
Db 497 IYVNCPEPIIDITFAITIRRLTYFFNLITPCVLIASMALGFTLPDPSGEKSLGVT 556
QY 304 ASLE-TMNRITAVTMEIS--VPVL 326
Db 557 ILLSLVFLMVAETMPATSDAVAL 582
RESULT 13
Q9RLA0 PRELIMINARY; PRT; 872 AA.
AC 09RLA0
ID 01-MAY-2000 (Tremblrel. 13, Created)
DE 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE DNA polymerase I.
GN POLA.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WTMINGTON;
RX MEDLINE=99416441; PubMed=10486973;
RA Anderson J.O., Anderson S.G.E.;
RT "Genome degradation is an ongoing process in Rickettsia."

RL MOL. Biol. Evol. 16:1178-1191(1999).
 DR EMBL: AJ238757; CAB56108.1; -
 DR HSSP: P00582; 1kFD.
 DR InterPro: IPR002421; 5_3_exonuclease.
 DR InterPro: IPR001098; DNA_pol.
 DR InterPro: IPR002298; DNA_pol.
 DR InterPro: IPR000513; Exo_N.1.
 DR InterPro: IPR003584; HNH.2.
 DR Pfam: PF01367; 5_3_exonuclease.1.
 DR Pfam: PF02739; 5_3_exonuclease.1.
 DR Pfam: PF00476; DNA_pol_A.1.
 DR PRINTS: PR00868; DNAPOL1.
 DR SMART: SM00475; 53EXOC.1.
 DR SMART: SM00279; Hnh2.1.
 DR SMART: SM00482; POLA.1.
 DR TIGRfams: TIGR00593; pola.1.
 DR PROSITE: PS00447; DNA_POLYMERASE.A.1.
 SQ SEQUENCE 872 AA; 99309 MW; BC2EBE3490879A08 CRC64;

Query Match 5.7%; Score 108; DB 2; Length 872;
 Best Local Similarity 20.4%; Pred. No. 6; 6; Mismatches 139; Indels 112; Gaps 15;

Matches 77; Conservative 50; Mismatches 139; Indels 112; Gaps 15;
 QY 2 NRIPLFTLFFSSVLTFFAVSADKIPGDESI-----TNIFGPRDRNESSPKH 48
 DB 383 NKSIIKITYSLKHLKLFVANSQSHKITAIEDLMQYTLASGLVQKNLFA-----ETLTMD 437
 QY 49 NILNNHITVSSSHLYIDMTFLCLSSHNLTN-----GACPTSSENPSSSVSGCTNIT 101
 DB 438 NINIESAKIVINISLY-KQTLALQKNKAFRLYREIDLPTECFILDKMERIGIKVDANLY 496
 QY 102 LQFTEK--RSLKRELOI-----KGKQLKFKSVNCPGSLFNSAHFNCNK 146
 DB 497 HOLSDPFGELIKIEEITLALSGTKFNIAQKOLSEILFKKQOLPSGNTLATSSISTK- 555
 QY 147 AASGASLYLYIPAGELKLPFGGIMDATLKLK-----VKRRYSEYGYTYTITITKL-- 198
 DB 556 -----AGILKLSSEGYHATILLRMRQTLKKNYTTSLPQIINNITRIHT 603
 QY 199 -----TDKGNIQIWLPOKSDARVDLNRPTGGGTYIGRNSVDMCFY--DGY---STNS 247
 DB 604 TELQSTTTLGRLSQEPNLQ-----NIPRSSDG-----NRIQAFIADGKILISADY 652
 QY 248 SLELEIRF-----ODNNPKSDGKFKYLRKINDTKE-----IAY 279
 DB 653 SOIELEIRIISHIANVDYLKCAFTNKEDIHTQACQIFNLQKHELTSEHRRKAKAINFGIY 712
 DB 280 TSLSLAGKSLPTNGTS 297
 DB 713 GISAFGLAKQALVNSNGTA 730
 RESULT 14
 ID 025579 PRELIMINARY; PRT: 2529 AA.
 AC 025579;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Toxin-like outer membrane protein.
 GN HP0922.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA *Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,

RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori."
 RL Nature 388:539-547(1997).
 DR EMBL: AE000602; AAD07969.1; -
 DR TIGR: HP0922;
 DR InterPro: IPR004311; Put_vacuole_cyttox.
 DR Pfam: PF03077; VacA2.3.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 2529 AA; 274561 MW; 440882E864472EC CRC64;

Query Match 5.7%; Score 108; DB 16; Length 2529;
 Best Local Similarity 22.5%; Pred. No. 25;
 Matches 84; Conservative 50; Mismatches 122; Indels 118; Gaps 22;

QY 85 TSENPSSSVSGSETNTLOFTEKRSII-----KRELOIKGYKQLFK---SVNCP 131
 DB 977 TSNFNATYQLGNTNFTL---SSQSLNFGDPTLQNNANITLGNKSQAAPKNSLTLQNN 1033
 QY 132 SGLTL-NSAHFNCNKNA--SGASLYLYPA-GELKNLPFGGIMDATLKLRYRRYSEY 187
 DB 1034 SNLSLDNQSVLNANNTSAPNQAASLNIYNGSQAFTNSLFPNG---GTLSLNSKILASN 1090
 QY 188 GYTTITITIKLTD-----KQNIQIWLPOKSDARVDLNRPTGGGTYIGRNSVD 236
 DB 1091 ASFSNNTTINLDDSVLASNTSLNANI-----NFGASQADP---CGNTIIFDASP 1140
 QY 237 MCFYDGYSTNSSL-----EIRQDNPK-----SDGKFLRKIN---DDT 274
 DB 1141 --FDSASSLNFNNLFRNGALNFRNGYPSLTALMSVSGFVLGNNDLNLSDINIFDNT 1198
 QY 275 KEIAYTSLLAGKSLPTNGTS-----LNIDAALEFN---NRRITAVTMEI 321
 DB 1199 KSVTY--NILNQKGLTIGSGANGYKILFYGMKQIONATYSDNNNTQMTSFLNPSQI 1256
 QY 322 -----SVPVL-----CWPRLQIDAKYENP-----EAGY--M 347
 DB 1257 IOESIKNGDLTTEVLNPNPSASNTIFNIAPELYNYQASKNPYGSYSDNQAGTYLT 1316
 QY 348 GINNTFTP-SSQF 360
 DB 1317 SNIKGLETPKGSQT 1330
 RESULT 15
 ID 08XAX2 PRELIMINARY; PRT: 304 AA.
 AC 08XAX2;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Putative adhesin, similar to FlmH protein.
 GN Z2206 OR PCS2107.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDU933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Heckert J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grodebeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
 RT Nature 409:529-533(2001).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-0157:H7 / RIMD 0509952;
 RX MEDLINE-21156231: Pubmed-11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 Iida T., Takami H., Honda T., Sasaki S., Ogasawara N., Yasunaga T.,
 Kuhara S., Shiba T., Hattori M., Shinagawa H.,
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 DR EMBL: AE005354; AAC56266.1; -
 DR EMBL: AP002557; BAB3530.1; -
 DR InterPro: IPR000259; Fimbrin.
 DR Pfam: PF00419; Fimbrin; 1.
 KW Complete proteome.
 SQ SEQUENCE 304 AA; 32041 MW; EED538023D95AFD5 CRC64;

Query Match 5.7%; Score 107; DB 16; Length 304;

Best Local Similarity 21.0%; Pred. No. 2;
 Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

QY 3 KIIF-ITLFFSSVLTFFAVSADKIPGDEST---TNIFFPRDRNESSPKHNI---LNH 54
 DB 8 KVLFGITLLMAGKVFAPFSCNVD--GSSSIGAGTTSYVNLDPVIOGQNLVVDLSQH 63
 QY 55 ITAYSESHLYDRMTFLCLSSH-NTLNGACPTSENSSSVSGETNITLQFTEKRSLIK 113
 DB 64 ISCWNDYGGWYD-----TDHINLVQG-----SAFAG----- 89
 QY 114 ELQIKGK-QLFKSVNCPGSLTNSAHFNCNKNASGASLYIIPAGELKNLPFGIWD 172
 DB 90 --SLQSYKGSLYMNNVYPPPLTTNTNVLIDIGDKTPMPLPKIYI-----TPVGAAG 140
 QY 173 ATK-----LRFKRRYSFTYGT-----YTINITIK-----LFDKGNIQIWLPOFK 212
 DB 141 VVIKAGEYIARIHMKIATLGSNGPRNFTWNIISNNSYVMPGTGCTYDSRNVTNLPDP 200
 QY 213 SDARVDLNLRPFGGTYIGRNSVDMCFYDGYSTNSSLIEIRFODNMPKSDGKFLRKIND 272
 DB 201 GSAEIPL-----GYVCS-SEQKLSFYLSGTTTDSARQV-FANTRAP-----D 239
 QY 273 DFKETATLSLLLAGK-----SLPTNGTSLNIADASLETWNRIYAVTMEI 321
 DB 240 ATKASGVGVSLMRNGKILATGENVSLGTIVNKSVPPLGSLATYGQTGNKVSAGTVQSV 296

Search completed: December 4, 2002, 17:28:40
 Time : 40 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2002, 17:21:12 : Search time 14 Seconds
(without alignments)
1069.497 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886

Sequence: 1 MNKILFIETLFSSVLTFFA.....EAGGYMGINVTFTPPSSQTL 361

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSPROT_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1558.5	82.6	360	1 CFAE_ECOLI	P25734 escherichia
2	120	6.4	2660	1 YEEJ_ECO57	O8X8V7 escherichia
3	114	6.0	650	1 SAGI_YEAST	P20840 saccharomyc
4	109	5.8	2358	1 YEEJ_ECOLI	P76347 escherichia
5	108	5.7	524	1 GAS3_YEAST	O03655 saccharomyc
6	107	5.7	304	1 YDEQ_ECOLI	P75588 escherichia
7	106	5.6	671	1 ALYS_ENTRA	P37110 enterococcu
8	99	5.2	1849	1 IGAA_HAEIN	P43386 haemophilus
9	98.5	5.2	608	1 YD93_METJA	O58788 methanococc
10	98	5.2	758	1 SP21_YEAST	P35209 saccharomyc
11	98	5.2	1157	1 NI13_YEAST	P36161 saccharomyc
12	95.5	5.1	713	1 CDG2_PAEMA	P31835 penicillium
13	95	5.0	413	1 RPIW_YEAST	P30775 saccharomyc
14	95	5.0	2334	1 WAPA_BACSU	O07833 bacillus su
15	94.5	5.0	959	1 NI00_YEAST	O02629 saccharomyc
16	94	5.0	471	1 LEU2_BUCRP	P48573 buchnera ap
17	93.5	5.0	412	1 THBG_SHEEP	P50450 ovis aries
18	93	4.9	1167	1 CUA_A_BACTU	P56956 bacillus th
19	92.5	4.9	320	1 CTF_CYACA	O96154 cyanidium c
20	92.5	4.9	544	1 AGM1_CANAL	O9P4V2 candida alb
21	92.5	4.9	1061	1 OAR_MYXXA	P36370 candida alb
22	92.5	4.9	2278	1 FABI_YEAST	P34756 saccharomyc
23	91.5	4.9	869	1 CFAE_ECOLI	P25733 escherichia
24	91.5	4.9	1783	1 Y468_MYCCE	O4460 mycoplasma
25	90.5	4.8	309	1 ELFB_CLOPE	P01586 clostridium
26	90.5	4.8	863	1 KCM4_XENLA	P36664 xenopus lae
27	90.5	4.8	1019	1 ENTK_HUMAN	P36073 homo sapien
28	90.5	4.8	1328	1 HUS2_SCHPO	O09811 schizosach
29	90.5	4.8	2193	1 POLG_HE7IM	O66479 human enter
30	89.5	4.7	1113	1 MGA2_YEAST	P40578 saccharomyc
31	89	4.7	465	1 ANT3_MOUSE	P32261 mus musculu
32	88.5	4.7	455	1 MURF_BUCAI	P57315 buchnera ap
33	88	4.7	1394	1 HAP_HAEIN	P45387 haemophilus

34	87.5	4.6	468	1 VGLC_HSYER	P12889 equine herp
35	87.5	4.6	792	1 UBPA_YEAST	P53874 saccharomyc
36	87	4.6	711	1 ERF2_SPKRA	O988Y2 shope fibro
37	87	4.6	1290	1 BXCL_CLOBO	P18640 clostridium
38	86.5	4.6	411	1 THBG_BOVIN	O9ct36 bos taurus
39	86.5	4.6	1002	1 SPHR_AMEPV	P29815 ansactra moo
40	86.5	4.6	1273	1 WEB1_YEAST	P38968 saccharomyc
41	86.5	4.6	1567	1 RML_DROME	O9y7H4 drosophila
42	86.5	4.6	1769	1 YK9_YEAST	P42945 saccharomyc
43	86	4.6	368	1 YBIL_YEAST	P38180 saccharomyc
44	86	4.6	493	1 FLIC_SALPA	P06178 salmoneila
45	86	4.6	828	1 MRKC_KLEPN	P21647 klebsiella

ALIGNMENTS

RESULT 1	CF AE_ECOLI	STANDARD:	PRT:	360 AA.
AC	P25734:			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	01-AUG-1992 (Rel. 23, Last annotation update)			
DE	CFA/I fimbrial subunit E (Colonization factor antigen I subunit E).			
GN	CFAE.			
OS	Escherichia coli.			
OG	Plasmid NRP513.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Enterotoxigenic;			
RX	MEDLINE=89330163; PubMed=2569152;			
RA	Hammers A.M., Pel H.J., Willshaw G.A., Kusters J.G.,			
RA	van der Zeijst B.A.M., Gaastra W.;			
RT	"The nucleotide sequence of the first two genes of the CFA/I fimbrial			
RT	operon of human enterotoxigenic Escherichia coli.";			
RL	Microb. Pathog. 6:297-309(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	MEDLINE=92329981; PubMed=1352712;			
RA	Jordi B.J.A.M., Willshaw G.A., van der Zeijst B.A.M., Gaastra W.;			
RT	"The complete nucleotide sequence of region I of the CFA/I fimbrial			
RT	operon of human enterotoxigenic Escherichia coli.";			
RL	DNA Seq. 2:257-263(1992).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
DR	EMBL: M55661; AAC4147.1; -			
KW	Antigen; Fimbria; Plasmid.			
SO	SEQUENCE 360 AA; 39903 MW; 691509B63A8E69CE CRC64;			
Query Match	82.6%; Score 1558.5; DB 1; Length 360;			
Best Local Similarity	81.2%; Pred. No. 4.7e-113;			
Matches	293; Conservative 31; Mismatches 36; Indels 1; Gaps 1;			
OY	1 MNKILFIETLFSSVLTFFPAVSADRI PGDESTINIFGPRDRNESSPKNINLNNHITAYSE 60			
DB	1 MNKILFIETLFSSVLTFFPAVSADRI PGDESTINIFGPRDRNESSPKNINLNNHITAYSE 60			
OY	61 SHTLVDRMTFLCSTSHNTLNGACPTSENPSSSVSGENITITQFEKSLTKREIQTIGY 120			
DB	61 SHTLVDRMTFLCSTSHNTLNGACPTSENPSSSVSGENITITQFEKSLTKREIQTIGY 120			
OY	121 KOLFRSVNCPSGTLTNSAHFNCNKRNAASGASLTYLIPAGELKNDLPFGIMDATILKRVK 180			
DB	121 KOLFRSVNCPSGTLTNSAHFNCNKRNAASGASLTYLIPAGELKNDLPFGIMDATILKRVK 180			

RL	MOL.	CELL.	BIOI.	9:3155-3165(1989).
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-S28bc;			
RX	MEDLINE=9000593; PubMed=267603;			
RA	Hauser K., Tanner W. ;			
RT	"Purification of the inducible alpha-agglutinin of S. cerevisiae and molecular cloning of the gene."			
RL	FEBB Lett. 255:290-294(1989).			
RN	(3)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-S28bc / FYJ679;			
RA	de Haan M., Smits P.H.M., Grievell L.A.;			
RL	submitted (May-1995) to the EMBL/Genebank/DBJ databases.			
RP	[4]			
RN	PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.			
RX	MEDLINE=96064684; PubMed=7592821;			
RA	Chen M.-H., Shen Z.-M., Robin S., Kahn P.C., Lipke P.N.;			
RT	*Structure of Saccharomyces cerevisiae alpha-agglutinin. Evidence for a yeast cell wall protein with multiple immunoglobulin-like domains with atypical disulfides.*;			
J.	Biol. Chem. 270:26168-26177(1995).			
CC	-1- FUNCTION: CELL SURFACE GLYCOPROTEIN PROMOTING CELL-CELL CONTACT TO FACILITATE MATING. SACCHAROMYCES CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-, RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR AGGREGATION DURING MATING.			
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POSSIBLE).			
CC	-1- INDUCTION: BY EXPOSITION TO PHEROMONE (A-FACTOR) SECRETED BY THE OPPOSITE MATING TYPE CELLS (TYPE A).			
CC	-1- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED.			
CC	-1- SIMILARITY: TO C.ALBCANS ALST.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isd-sib.ch/announce/ or send an email to license@sib-stb.ch).			
CC	-----			
DR	EMBL; M28164; AAA34417.1; -			
DR	EMBL; X16861; CA934752.1; -			
DR	EMBL; X87611; CAA60926.1; -			
DR	EMBL; 249504; CAA89526.1; -			
DR	PIR; S22835; S22835.			
DR	SGD; S0003764; SAG1.			
RZ	Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat.			
FT	SIGNAL	1	19	
FT	CHAIN	20	650	
FT	DOMAIN	278	350	
FT				ALPHA-AGGLUTININ.
FT				ACIDIC REGION, PROBABLY IMPORTANT IN
FT	DONAIN	339	423	BINDING TO AGGLUTININS OF TYPE A CELLS.
FT	REPEAT	339	378	2 x 40 AA TANDEM REPEATS.
FT	REPEAT	384	423	1.
FT	DISULFID	97	114	2.
FT	CARBOHYD	202	300	
FT	CARBOHYD	79	79	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	109	109	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	135	135	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	248	248	O-LINKED (GLCNAC . .) .
FT	CARBOHYD	282	282	O-LINKED (MAN . .) .
FT	CARBOHYD	289	289	O-LINKED (MAN . .) .
FT	CARBOHYD	299	299	O-LINKED (MAN . .) .
FT	CARBOHYD	303	303	O-LINKED (MAN . .) .
FT	CARBOHYD	306	306	N-LINKED (GLCNAC . .) .
FT	CARBOHYD	307	307	O-LINKED (MAN . .) .
FT	CARBOHYD	308	308	O-LINKED (MAN . .) .
FT	CARBOHYD	311	311	O-LINKED (MAN . .) .
FT	CARBOHYD	314	314	O-LINKED (MAN . .) .
FT	CARBOHYD	315	315	O-LINKED (MAN . .) .
FT	CARBOHYD	316	316	O-LINKED (MAN . .) .

Query Match	Best Local Similarity	6.0%;	Score 114;	DB 1;	Length 650;
Matches	98;	Conservative	52;	Mismatches 147;	Indels 172; Gaps 23;
QY 4 ILFIITLFFSSVL-----PFPA-----VSADKIPG-----DESTINIFGRDNE---	43				
DB 8 ILMLFSLALASXININDITFSSMLETTPITANKOPDQGWATPFDSTADASSIREGDEFTL	67				
QY 44 SSP---KHNIILNNHTA-----YSESHLYDRMTFLC-----LSSHNTLN	80				
DB 68 SMPHYRIKLKLNSSQFATISLADGEAFRCYVSGQAAYIYENTFTTCTAQNDSIREGDEFTL	127				
QY 81 GACPISENPPSSSVSGENITITLOTFEKSILRIEIQIKYKOLLKRSYVC-PSGLTNSA	139				
DB 128 GSITPSLNFSPDGSSEYEEL-----ENAKFFSGPMLVLKGNQMSDVVNFDAAFTEYVF	182				
QY 140 HENCNKNAASGASLVL--YIPAGELKNLPFGIMPATLTKRKRRYSSELYGTITINITTK	197				
DB 183 HSGRSTGYGSEFSYHLGMCPNGYF---LGG-----TEKI-----DYDSSNNVLD-	225				
QY 198 LTKGNIGI-----WLPQFKSDARYDLMLRPTGGGTYIGRN---SVDCCFYDG--	242				
DB 226 -DCSSVQYVSSNDPMDWMFPPSYNDTNAV-----TCGNSLMITLDEKIVDGEWL	275				
QY 243 -----YSTNSSLEF--IRFODNN--PKSDGKF	265				
DB 276 WYNALQSLPANVNTIDHLEFYQTCLDITANTTYATQFSTREFIITYQGRNLTGTASAKSS	335				
QY 266 YLRKINDDTKEI--AYILS-----LLAGKSLTPTNGSLNINADAA	304				
DB 336 FISTTTTDLTSLINTSAVSTGSIISVETGNGRTTSEVISHWVTTSTKLSPATTSTLTAQTS	395				
QY 305 SLEETMNN-----RITAVTPEISVPVLCMPGRGL	332				
DB 396 IYSTDSNITVGDTHITTSSEVTSIDVETISKEHTASTYVAAPLTSTTGNTGAM	444				
RESULT 4					
YEEB_ECOLI					
ID YEEB_ECOLI	STANDARD:				PRT: 2358 AA.
AC P76347; P94750;					
DT 15-JUN-2002 (Rel. 41, Created)					
DT 15-JUN-2002 (Rel. 41, Last sequence update)					
DT 15-JUN-2002 (Rel. 41, Last annotation update)					
DE Hypothetical protein yeeJ.					
GN YEEB OR B1978.					
OS Escherichia coli.					
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;					
OC Escherichia.					

```

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Ma B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97251358; PubMed-9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Sivasubraman S., Tagami H.,
RA Takeuchi J., Takemoto K., Wada C., Yamamoto Y., Horiiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -1- SIMILARITY: CONTAINS 13 BIG-1 DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: AEO00289; AAC75042.1; ALT_INIT.
DR EMBL: D90837; BAA15800.1; -.
DR EMBL: D90836; BAA15799.1; ALT_INIT.
DR EcoGene: EG13378; Yeel.
DR InterPro: IPR003344; Big_1.
DR InterPro: IPR003535; Intimin.
DR InterPro: IPR002482; LysM.
DR InterPro: IPR000601; PKD_domain.
DR Pfam: PF02369; Big_1; 13.
DR PRINTS: PRO1369; Intimin.
DR SMART: SM00257; LysM; 1.
DR SMART: SM00089; PKD; 4.
KM Hypothetical protein; Repeat; Complete proteome.
DR DOMAIN 738 834
RT BIG-1 1.
FT 840 931
DR DOMAIN 932 1033
RT BIG-1 2.
FT 1042 1137
DR DOMAIN 1146 1237
RT BIG-1 3.
FT 1246 1350
DR DOMAIN 1351 1448
RT BIG-1 4.
FT 1449 1533
DR DOMAIN 1534 1635
RT BIG-1 5.
FT 1661 1754
DR DOMAIN 1763 1853
RT BIG-1 6.
FT 1855 1950
DR DOMAIN 1952 2053
RT BIG-1 7.
FT 2054 2148
DR DOMAIN 2149 2248
RT BIG-1 8.
FT 2249 2348
DR DOMAIN 2349 2448
RT BIG-1 9.
FT 2449 2548
DR DOMAIN 2549 2648
RT BIG-1 10.
FT 2649 2748
DR DOMAIN 2749 2848
RT BIG-1 11.
FT 2849 2948
DR DOMAIN 2949 3048
RT BIG-1 12.
FT 3049 3148
DR DOMAIN 3149 3248
RT BIG-1 13.
FT 3249 3348
DR DOMAIN 3349 3448
RT BIG-1 14.
FT 3449 3548
DR DOMAIN 3549 3648
RT BIG-1 15.
FT 3649 3748
DR DOMAIN 3749 3848
RT BIG-1 16.
FT 3849 3948
DR DOMAIN 3949 4048
RT BIG-1 17.
FT 4049 4148
DR DOMAIN 4149 4248
RT BIG-1 18.
FT 4249 4348
DR DOMAIN 4349 4448
RT BIG-1 19.
FT 4449 4548
DR DOMAIN 4549 4648
RT BIG-1 20.
FT 4649 4748
DR DOMAIN 4749 4848
RT BIG-1 21.
FT 4849 4948
DR DOMAIN 4949 5048
RT BIG-1 22.
FT 5049 5148
DR DOMAIN 5149 5248
RT BIG-1 23.
FT 5249 5348
DR DOMAIN 5349 5448
RT BIG-1 24.
FT 5449 5548
DR DOMAIN 5549 5648
RT BIG-1 25.
FT 5649 5748
DR DOMAIN 5749 5848
RT BIG-1 26.
FT 5849 5948
DR DOMAIN 5949 6048
RT BIG-1 27.
FT 6049 6148
DR DOMAIN 6149 6248
RT BIG-1 28.
FT 6249 6348
DR DOMAIN 6349 6448
RT BIG-1 29.
FT 6449 6548
DR DOMAIN 6549 6648
RT BIG-1 30.
FT 6649 6748
DR DOMAIN 6749 6848
RT BIG-1 31.
FT 6849 6948
DR DOMAIN 6949 7048
RT BIG-1 32.
FT 7049 7148
DR DOMAIN 7149 7248
RT BIG-1 33.
FT 7249 7348
DR DOMAIN 7349 7448
RT BIG-1 34.
FT 7449 7548
DR DOMAIN 7549 7648
RT BIG-1 35.
FT 7649 7748
DR DOMAIN 7749 7848
RT BIG-1 36.
FT 7849 7948
DR DOMAIN 7949 8048
RT BIG-1 37.
FT 8049 8148
DR DOMAIN 8149 8248
RT BIG-1 38.
FT 8249 8348
DR DOMAIN 8349 8448
RT BIG-1 39.
FT 8449 8548
DR DOMAIN 8549 8648
RT BIG-1 40.
FT 8649 8748
DR DOMAIN 8749 8848
RT BIG-1 41.
FT 8849 8948
DR DOMAIN 8949 9048
RT BIG-1 42.
FT 9049 9148
DR DOMAIN 9149 9248
RT BIG-1 43.
FT 9249 9348
DR DOMAIN 9349 9448
RT BIG-1 44.
FT 9449 9548
DR DOMAIN 9549 9648
RT BIG-1 45.
FT 9649 9748
DR DOMAIN 9749 9848
RT BIG-1 46.
FT 9849 9948
DR DOMAIN 9949 10048
RT BIG-1 47.
FT 10049 10148
DR DOMAIN 10149 10248
RT BIG-1 48.
FT 10249 10348
DR DOMAIN 10349 10448
RT BIG-1 49.
FT 10449 10548
DR DOMAIN 10549 10648
RT BIG-1 50.
FT 10649 10748
DR DOMAIN 10749 10848
RT BIG-1 51.
FT 10849 10948
DR DOMAIN 10949 11048
RT BIG-1 52.
FT 11049 11148
DR DOMAIN 11149 11248
RT BIG-1 53.
FT 11249 11348
DR DOMAIN 11349 11448
RT BIG-1 54.
FT 11449 11548
DR DOMAIN 11549 11648
RT BIG-1 55.
FT 11649 11748
DR DOMAIN 11749 11848
RT BIG-1 56.
FT 11849 11948
DR DOMAIN 11949 12048
RT BIG-1 57.
FT 12049 12148
DR DOMAIN 12149 12248
RT BIG-1 58.
FT 12249 12348
DR DOMAIN 12349 12448
RT BIG-1 59.
FT 12449 12548
DR DOMAIN 12549 12648
RT BIG-1 60.
FT 12649 12748
DR DOMAIN 12749 12848
RT BIG-1 61.
FT 12849 12948
DR DOMAIN 12949 13048
RT BIG-1 62.
FT 13049 13148
DR DOMAIN 13149 13248
RT BIG-1 63.
FT 13249 13348
DR DOMAIN 13349 13448
RT BIG-1 64.
FT 13449 13548
DR DOMAIN 13549 13648
RT BIG-1 65.
FT 13649 13748
DR DOMAIN 13749 13848
RT BIG-1 66.
FT 13849 13948
DR DOMAIN 13949 14048
RT BIG-1 67.
FT 14049 14148
DR DOMAIN 14149 14248
RT BIG-1 68.
FT 14249 14348
DR DOMAIN 14349 14448
RT BIG-1 69.
FT 14449 14548
DR DOMAIN 14549 14648
RT BIG-1 70.
FT 14649 14748
DR DOMAIN 14749 14848
RT BIG-1 71.
FT 14849 14948
DR DOMAIN 14949 15048
RT BIG-1 72.
FT 15049 15148
DR DOMAIN 15149 15248
RT BIG-1 73.
FT 15249 15348
DR DOMAIN 15349 15448
RT BIG-1 74.
FT 15449 15548
DR DOMAIN 15549 15648
RT BIG-1 75.
FT 15649 15748
DR DOMAIN 15749 15848
RT BIG-1 76.
FT 15849 15948
DR DOMAIN 15949 16048
RT BIG-1 77.
FT 16049 16148
DR DOMAIN 16149 16248
RT BIG-1 78.
FT 16249 16348
DR DOMAIN 16349 16448
RT BIG-1 79.
FT 16449 16548
DR DOMAIN 16549 16648
RT BIG-1 80.
FT 16649 16748
DR DOMAIN 16749 16848
RT BIG-1 81.
FT 16849 16948
DR DOMAIN 16949 17048
RT BIG-1 82.
FT 17049 17148
DR DOMAIN 17149 17248
RT BIG-1 83.
FT 17249 17348
DR DOMAIN 17349 17448
RT BIG-1 84.
FT 17449 17548
DR DOMAIN 17549 17648
RT BIG-1 85.
FT 17649 17748
DR DOMAIN 17749 17848
RT BIG-1 86.
FT 17849 17948
DR DOMAIN 17949 18048
RT BIG-1 87.
FT 18049 18148
DR DOMAIN 18149 18248
RT BIG-1 88.
FT 18249 18348
DR DOMAIN 18349 18448
RT BIG-1 89.
FT 18449 18548
DR DOMAIN 18549 18648
RT BIG-1 90.
FT 18649 18748
DR DOMAIN 18749 18848
RT BIG-1 91.
FT 18849 18948
DR DOMAIN 18949 19048
RT BIG-1 92.
FT 19049 19148
DR DOMAIN 19149 19248
RT BIG-1 93.
FT 19249 19348
DR DOMAIN 19349 19448
RT BIG-1 94.
FT 19449 19548
DR DOMAIN 19549 19648
RT BIG-1 95.
FT 19649 19748
DR DOMAIN 19749 19848
RT BIG-1 96.
FT 19849 19948
DR DOMAIN 19949 20048
RT BIG-1 97.
FT 20049 20148
DR DOMAIN 20149 20248
RT BIG-1 98.
FT 20249 20348
DR DOMAIN 20349 20448
RT BIG-1 99.
FT 20449 20548
DR DOMAIN 20549 20648
RT BIG-1 100.
FT 20649 20748
DR DOMAIN 20749 20848
RT BIG-1 101.
FT 20849 20948
DR DOMAIN 20949 21048
RT BIG-1 102.
FT 21049 21148
DR DOMAIN 21149 21248
RT BIG-1 103.
FT 21249 21348
DR DOMAIN 21349 21448
RT BIG-1 104.
FT 21449 21548
DR DOMAIN 21549 21648
RT BIG-1 105.
FT 21649 21748
DR DOMAIN 21749 21848
RT BIG-1 106.
FT 21849 21948
DR DOMAIN 21949 22048
RT BIG-1 107.
FT 22049 22148
DR DOMAIN 22149 22248
RT BIG-1 108.
FT 22249 22348
DR DOMAIN 22349 22448
RT BIG-1 109.
FT 22449 22548
DR DOMAIN 22549 22648
RT BIG-1 110.
FT 22649 22748
DR DOMAIN 22749 22848
RT BIG-1 111.
FT 22849 22948
DR DOMAIN 22949 23048
RT BIG-1 112.
FT 23049 23148
DR DOMAIN 23149 23248
RT BIG-1 113.
FT 23249 23348
DR DOMAIN 23349 23448
RT BIG-1 114.
FT 23449 23548
DR DOMAIN 23549 23648
RT BIG-1 115.
FT 23649 23748
DR DOMAIN 23749 23848
RT BIG-1 116.
FT 23849 23948
DR DOMAIN 23949 24048
RT BIG-1 117.
FT 24049 24148
DR DOMAIN 24149 24248
RT BIG-1 118.
FT 24249 24348
DR DOMAIN 24349 24448
RT BIG-1 119.
FT 24449 24548
DR DOMAIN 24549 24648
RT BIG-1 120.
FT 24649 24748
DR DOMAIN 24749 24848
RT BIG-1 121.
FT 24849 24948
DR DOMAIN 24949 25048
RT BIG-1 122.
FT 25049 25148
DR DOMAIN 25149 25248
RT BIG-1 123.
FT 25249 25348
DR DOMAIN 25349 25448
RT BIG-1 124.
FT 25449 25548
DR DOMAIN 25549 25648
RT BIG-1 125.
FT 25649 25748
DR DOMAIN 25749 25848
RT BIG-1 126.
FT 25849 25948
DR DOMAIN 25949 26048
RT BIG-1 127.
FT 26049 26148
DR DOMAIN 26149 26248
RT BIG-1 128.
FT 26249 26348
DR DOMAIN 26349 26448
RT BIG-1 129.
FT 26449 26548
DR DOMAIN 26549 26648
RT BIG-1 130.
FT 26649 26748
DR DOMAIN 26749 26848
RT BIG-1 131.
FT 26849 26948
DR DOMAIN 26949 27048
RT BIG-1 132.
FT 27049 27148
DR DOMAIN 27149 27248
RT BIG-1 133.
FT 27249 27348
DR DOMAIN 27349 27448
RT BIG-1 134.
FT 27449 27548
DR DOMAIN 27549 27648
RT BIG-1 135.
FT 27649 27748
DR DOMAIN 27749 27848
RT BIG-1 136.
FT 27849 27948
DR DOMAIN 27949 28048
RT BIG-1 137.
FT 28049 28148
DR DOMAIN 28149 28248
RT BIG-1 138.
FT 28249 28348
DR DOMAIN 28349 28448
RT BIG-1 139.
FT 28449 28548
DR DOMAIN 28549 28648
RT BIG-1 140.
FT 28649 28748
DR DOMAIN 28749 28848
RT BIG-1 141.
FT 28849 28948
DR DOMAIN 28949 29048
RT BIG-1 142.
FT 29049 29148
DR DOMAIN 29149 29248
RT BIG-1 143.
FT 29249 29348
DR DOMAIN 29349 29448
RT BIG-1 144.
FT 29449 29548
DR DOMAIN 29549 29648
RT BIG-1 145.
FT 29649 29748
DR DOMAIN 29749 29848
RT BIG-1 146.
FT 29849 29948
DR DOMAIN 29949 30048
RT BIG-1 147.
FT 30049 30148
DR DOMAIN 30149 30248
RT BIG-1 148.
FT 30249 30348
DR DOMAIN 30349 30448
RT BIG-1 149.
FT 30449 30548
DR DOMAIN 30549 30648
RT BIG-1 150.
FT 30649 30748
DR DOMAIN 30749 30848
RT BIG-1 151.
FT 30849 30948
DR DOMAIN 30949 31048
RT BIG-1 152.
FT 31049 31148
DR DOMAIN 31149 31248
RT BIG-1 153.
FT 31249 31348
DR DOMAIN 31349 31448
RT BIG-1 154.
FT 31449 31548
DR DOMAIN 31549 31648
RT BIG-1 155.
FT 31649 31748
DR DOMAIN 31749 31848
RT BIG-1 156.
FT 31849 31948
DR DOMAIN 31949 32048
RT BIG-1 157.
FT 32049 32148
DR DOMAIN 32149 32248
RT BIG-1 158.
FT 32249 32348
DR DOMAIN 32349 32448
RT BIG-1 159.
FT 32449 32548
DR DOMAIN 32549 32648
RT BIG-1 160.
FT 32649 32748
DR DOMAIN 32749 32848
RT BIG-1 161.
FT 32849 32948
DR DOMAIN 32949 33048
RT BIG-1 162.
FT 33049 33148
DR DOMAIN 33149 33248
RT BIG-1 163.
FT 33249 33348
DR DOMAIN 33349 33448
RT BIG-1 164.
FT 33449 33548
DR DOMAIN 33549 33648
RT BIG-1 165.
FT 33649 33748
DR DOMAIN 33749 33848
RT BIG-1 166.
FT 33849 33948
DR DOMAIN 33949 34048
RT BIG-1 167.
FT 34049 34148
DR DOMAIN 34149 34248
RT BIG-1 168.
FT 34249 34348
DR DOMAIN 34349 34448
RT BIG-1 169.
FT 34449 34548
DR DOMAIN 34549 34648
RT BIG-1 170.
FT 34649 34748
DR DOMAIN 34749 34848
RT BIG-1 171.
FT 34849 34948
DR DOMAIN 34949 35048
RT BIG-1 172.
FT 35049 35148
DR DOMAIN 35149 35248
RT BIG-1 173.
FT 35249 35348
DR DOMAIN 35349 35448
RT BIG-1 174.
FT 35449 35548
DR DOMAIN 35549 35648
RT BIG-1 175.
FT 35649 35748
DR DOMAIN 35749 35848
RT BIG-1 176.
FT 35849 35948
DR DOMAIN 35949 36048
RT BIG-1 177.
FT 36049 36148
DR DOMAIN 36149 36248
RT BIG-1 178.
FT 36249 36348
DR DOMAIN 36349 36448
RT BIG-1 179.
FT 36449 36548
DR DOMAIN 36549 36648
RT BIG-1 180.
FT 36649 36748
DR DOMAIN 36749 36848
RT BIG-1 181.
FT 36849 36948
DR DOMAIN 36949 37048
RT BIG-1 182.
FT 37049 37148
DR DOMAIN 37149 37248
RT BIG-1 183.
FT 37249 37348
DR DOMAIN 37349 37448
RT BIG-1 184.
FT 37449 37548
DR DOMAIN 37549 37648
RT BIG-1 185.
FT 37649 37748
DR DOMAIN 37749 37848
RT BIG-1 186.
FT 37849 37948
DR DOMAIN 37949 38048
RT BIG-1 187.
FT 38049 38148
DR DOMAIN 38149 38248
RT BIG-1 188.
FT 38249 38348
DR DOMAIN 38349 38448
RT BIG-1 189.
FT 38449 38548
DR DOMAIN 38549 38648
RT BIG-1 190.
FT 38649 38748
DR DOMAIN 38749 38848
RT BIG-1 191.
FT 38849 38948
DR DOMAIN 38949 39048
RT BIG-1 192.
FT 39049 39148
DR DOMAIN 39149 39248
RT BIG-1 193.
FT 39249 39348
DR DOMAIN 39349 39448
RT BIG-1 194.
FT 39449 39548
DR DOMAIN 39549 39648
RT BIG-1 195.
FT 39649 39748
DR DOMAIN 39749 39848
RT BIG-1 196.
FT 39849 39948
DR DOMAIN 39949 40048
RT BIG-1 197.
FT 40049 40148
DR DOMAIN 40149 40248
RT BIG-1 198.
FT 40249 40348
DR DOMAIN 40349 40448
RT BIG-1 199.
FT 40449 40548
DR DOMAIN 40549 40648
RT BIG-1 200.
FT 40649 40748
DR DOMAIN 40749 40848
RT BIG-1 201.
FT 40849 40948
DR DOMAIN 40949 41048
RT BIG-1 202.
FT 41049 41148
DR DOMAIN 41149 41248
RT BIG-1 203.
FT 41249 41348
DR DOMAIN 41349 41448
RT BIG-1 204.
FT 41449 41548
DR DOMAIN 41549 41648
RT BIG-1 205.
FT 41649 41748
DR DOMAIN 41749 41848
RT BIG-1 206.
FT 41849 41948
DR DOMAIN 41949 42048
RT BIG-1 207.
FT 42049 42148
DR DOMAIN 42149 42248
RT BIG-1 208.
FT 42249 42348
DR DOMAIN 42349 42448
RT BIG-1 209.
FT 42449 42548
DR DOMAIN 42549 42648
RT BIG-1 210.
FT 42649 42748
DR DOMAIN 42749 42848
RT BIG-1 211.
FT 42849 42948
DR DOMAIN 42949 43048
RT BIG-1 212.
FT 43049 43148
DR DOMAIN 43149 43248
RT BIG-1 213.
FT 43249 43348
DR DOMAIN 43349 43448
RT BIG-1 214.
FT 43449 43548
DR DOMAIN 43549 43648
RT BIG-1 215.
FT 43649 43748
DR DOMAIN 43749 43848
RT BIG-1 216.
FT 43849 43948
DR DOMAIN 43949 44048
RT BIG-1 217.
FT 44049 44148
DR DOMAIN 44149 44248
RT BIG-1 218.
FT 44249 44348
DR DOMAIN 44349 44448
RT BIG-1 219.
FT 44449 44548
DR DOMAIN 44549 44648
RT BIG-1 220.
FT 44649 44748
DR DOMAIN 44749 44848
RT BIG-1 221.
FT 44849 44948
DR DOMAIN 44949 45048
RT BIG-1 222.
FT 45049 45148
DR DOMAIN 45149 45248
RT BIG-1 223.
FT 45249 45348
DR DOMAIN 45349 45448
RT BIG-1 224.
FT 45449 45548
DR DOMAIN 45549 45648
RT BIG-1 225.
FT 45649 45748
DR DOMAIN 45749 45848
RT BIG-1 226.
FT 45849 45948
DR DOMAIN 45949 46048
RT BIG-1 227.
FT 46049 46148
DR DOMAIN 46149 46248
RT BIG-1 228.
FT 46249 46348
DR DOMAIN 46349 46448
RT BIG-1 229.
FT 46449 46548
DR DOMAIN 46549 46648
RT BIG-1 230.
FT 46649 46748
DR DOMAIN 46749 46848
RT BIG-1 231.
FT 46849 46948
DR DOMAIN 46949 47048
RT BIG-1 232.
FT 47049 47148
DR DOMAIN 47149 47248
RT BIG-1 233.
FT 47249 47348
DR DOMAIN 47349 47448
RT BIG-1 234.
FT 47449 47548
DR DOMAIN 47549 47648
RT BIG-1 235.
FT 47649 47748
DR DOMAIN 47749 47848
RT BIG-1 236.
FT 47849 47948
DR DOMAIN 47949 48048
RT BIG-1 237.
FT 48049 48148
DR DOMAIN 48149 48248
RT BIG-1 238.
FT 48249 48348
DR DOMAIN 48349 48448
RT BIG-1 239.
FT 48449 48548
DR DOMAIN 48549 48648
RT BIG-1 240.
FT 48649 48748
DR DOMAIN 48749 48848
RT BIG-1 241.
FT 48849 48948
DR DOMAIN 48949 49048
RT BIG-1 242.
FT 49049 49148
DR DOMAIN 49149 49248
RT BIG-1 243.
FT 49249 49348
DR DOMAIN 49349 49448
RT BIG-1 244.
FT 49449 49548
DR DOMAIN 49549 49648
RT BIG-1 245.
FT 49649 49748
DR DOMAIN 49749 49848
RT BIG-1 246.
FT 49849 49948
DR DOMAIN 49949 50048
RT BIG-1 247.
FT 50049 50148
DR DOMAIN 50149 50248
RT BIG-1 248.
FT 50249 50348
DR DOMAIN 50349 50448
RT BIG-1 249.
FT 50449 50548
DR DOMAIN 50549 50648
RT BIG-1 250.
FT 50649 50748
DR DOMAIN 50749 50848
RT BIG-1 251.
FT 50849 50948
DR DOMAIN 50949 51048
RT BIG-1 252.
FT 51049 51148
DR DOMAIN 51149 51248
RT BIG-1 253.
FT 51249 51348
DR DOMAIN 51349 51448
RT BIG-1 254.
FT 51449 51548
DR DOMAIN 51549 51648
RT BIG-1 255.
FT 51649 51748
DR DOMAIN 51749 51848
RT BIG-1 256.
FT 51849 51948
DR DOMAIN 51949 52048
RT BIG-1 257.
FT 52049 52148
DR DOMAIN 52149 52248
RT BIG-1 258.
FT 52249 52348
DR DOMAIN 52349 52448
RT BIG-1 259.
FT 52449 52548
DR DOMAIN 52549 52648
RT BIG-1 260.
FT 52649 52748
DR DOMAIN 52749 52848
RT BIG-1 261.
FT 52849 52948
DR DOMAIN 52949 53048
RT BIG-1 262.
FT 53049 53148
DR DOMAIN 53149 53248
RT BIG-1 263.
FT 53249 53348
DR DOMAIN 53349 53448
RT BIG-1 264.
FT 53449 53548
DR DOMAIN 53549 53648
RT BIG-1 265.
FT 53649 53748
DR DOMAIN 53749 53848
RT BIG-1 266.
FT 53849 53948
DR DOMAIN 53949 54048
RT BIG-1 267.
FT 54049 54148
DR DOMAIN 54149 54248
RT BIG-1 268.
FT 54249 54348
DR DOMAIN 54349 54448
RT BIG-1 269.
FT 54449 54548
DR DOMAIN 54549 54648
RT BIG-1 270.
FT 54649 54748
DR DOMAIN 54749 54848
RT BIG-1 271.
FT 54849 54948
DR DOMAIN 54949 55048
RT BIG-1 272.
FT 55049 55148
DR DOMAIN 55149 55248
RT BIG-1 273.
FT 55249 55348
DR DOMAIN 55349 55448
RT BIG-1 274.
FT 55449 55548
DR DOMAIN 55549 55648
RT BIG-1 275.
FT 55649 55748
DR DOMAIN 55749 55848
RT BIG-1 276.
FT 55849 55948
DR DOMAIN 55949 56048
RT BIG-1 277.
FT 56049 56148
DR DOMAIN 56149 56248
RT BIG-1 278.
FT 56249 56348
DR DOMAIN 56349 56448
RT BIG-1 279.
FT 56449 56548
DR DOMAIN 56549 56648
RT BIG-1 280.
FT 56649 56748
DR DOMAIN 56749 56848
RT BIG-1 281.
FT 56849 56948
DR DOMAIN 56949 57048
RT BIG-1 282.
FT 57049 57148
DR DOMAIN 57149 57248
RT BIG-1 283.
FT 57249 57348
DR DOMAIN 57349 57448
RT BIG-1 284.
FT 57449 57548
DR DOMAIN 57549 57648
RT BIG-1 285.
FT 57649 57748
DR DOMAIN 57749 57848
RT BIG-1 286.
FT 57849 57948
DR DOMAIN 57949 58048
RT BIG-1 287.
FT 58049 58148
DR DOMAIN 58149 58248
RT BIG-1 288.
FT 58249 58348
DR DOMAIN 58349 58448
RT BIG-1 289.
FT 58449 58548
DR DOMAIN 58549 58648
RT BIG-1 290.
FT 58649 58748
DR DOMAIN 58749 58848
RT BIG-1 291.
FT 58849 58948
DR DOMAIN 58949 59048
RT BIG-1 292.
FT 59049 59148
DR DOMAIN 59149 59248
RT BIG-1 293.
FT 59249 59348
DR DOMAIN 59349 59448
RT BIG-1 294.
FT 59449 59548
DR DOMAIN 59549 59648
RT BIG-1 295.
FT 59649 59748
DR DOMAIN 59749 59848
RT BIG-1 296.
FT 59849 59948
DR DOMAIN 59949 60048
RT BIG-1 297.
FT 60049 60148
DR DOMAIN 60149 60248
RT BIG-1 298.
FT 60249 60348

```



```

AC Q58788;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1393.
GN MJ1393.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococci: Methanococcales:
OC Methanocaldococcaceae: Methanocaldococcus.
OX NCBI_TaxID:2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE:96337999; PubMed:8668087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerevage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weisslock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Meldrum J.F., Fuhmann J.L., Nguyen D.,
RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: TO M.JANNASCHII MJ1394 AND A.FULGIDUS AF2028.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL: U67579; AAB94403.1; -.
CC DR TIGR: MJ1393; -.
CC -KM Hypothetical protein; Transmembrane; Complete proteome.
CC FT TRANSMEM 4 24 POTENTIAL.
CC SO SEQUENCE 608 AA; 66768 MW; 010FAFLC29FC873C CR664;
Query Match 5.2%; Score 98.5; DB 1; Length 608;
Best Local Similarity 20.3%; Pred. No. 3.5;
Matches 85; Conservative 68; Mismatches 119; Indels 147; Gaps 25;
QY 1 MKKILF---ITLFPSSVLTFAVSADKIP-----GDSINIRPROR 41
QY 1 MKKILFMAIIMSLIFGTVFGY---GDNGLVYVAYEYKYNITGNTGDIVSSTI----- 52
QY 42 NSSPKHNLNHNHTAYSSHTLYDMFTLCSSHNT-----LNG-----ACPTSE 87
QY 53 -ESTIGYIYNN--TGTINDTLIDYVAVANVSNNTTGEVYVNGTPKVPFISSAPAYT 109
QY 88 N-PSSSS-----VSGETNITLQFTKRSLLKRELOIKGYKQLLFKSVNCPGSLTLN-SA 139
QY 110 NLPNATVYIHPILPNNYSVIAIKFAIDKSITGVPILINE-----TYSDTKIPSPRLSNWGV 165
QY 140 HFNCKNNAASGSLYITAGE-----LKNLPFGINDATLKLKVKRRISTYGT 189
QY 166 YLNISSNVA---LPAITDPVSVIMTKYLSNDP-----NNYS 200
QY 190 YT---INIKITLKGNIQIWM---LPOFK-----SDARYDLNLRPTGGGT 228
QY 201 DPMNPLNITGALANESITLMDGPYFLPGYNDLSLTWTGAVINTKNAITITINI--TGNNT 258
QY 229 YIGRNSVDMCFDYSTNSSLEIRFQDNNPKSDGKFLYRKINDTKELAYT---LSL 283
QY 259 YTNRTGTLMKY--GEAV-----IIEEFNGTKSGTKI-----EGIVATGYGVGSA 300
QY 284 LLAGKSLTPTNG-----TSLNIADAASLETNMRITAYTMEPI--SVPYLCMPGRLQ 334
QY 301 TREGPLNASSGKYELIWEBSANVSNKAS--SYFNLTHVTIWAVNGSNPVIIDPFENITL 357

```

```

RESULT 10
SP21_YEAST
ID SP21_YEAST STANDARD; PRT; 758 AA.
AC P35209;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SP21 protein.
GN SP21 OR YMR179W OR YMR8010.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:94186069; PubMed:8138180;
RA Natsoulis G., Winston F., Boeke J.D.;
RT "The SP21 and SP22 genes of Saccharomyces cerevisiae."
RL Genetics 136:93-105(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR NORMAL TRANSCRIPTION AT A NUMBER OF LOC1 IN
CC YEAST.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL: L24436; AAA35078.1; -.
CC DR EMBL: 249808; CAA89912.1; -.
CC DR PIR: S47866; S47866.
CC DR TRANSPAC: T04376; -.
CC DR SGD: S0004791; SEP21.
CC DR DOMAIN 127 144 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 672 682 ASP/GLU-RICH (ACIDIC).
CC SO SEQUENCE 758 AA; 84697 MW; 7DB3FCF7EE96705 CR664;
Query Match 5.2%; Score 98; DB 1; Length 758;
Best Local Similarity 21.9%; Pred. No. 5.1;
Matches 61; Conservative 34; Mismatches 95; Indels 88; Gaps 12;
QY 42 NSSPKHNLNHNHTAYSSHTLYDMFT-----ICSSHNTLNGACPTSENPSSSVS 95
QY 166 NISKKGRVANNQI---PEETLEVKALRTKYITNTSGNNTTNSRISCLDPSSL--- 218
QY 96 GETNITLQFTKRSLLKRELOIKGYKQLLFKSVNCPGSLTLSAHFNCKNNAASGSLYL 155
QY 219 -PSALPTPTPSQSGLFNQIK-----NSRNATITIN-----NNSGT----- 257
QY 156 YIPAGEELKNLPFGIWDATLKLKVKRRYSEYTYGTYINITIKLTDKNIQIWLPOFKSDA 215
QY 258 ---VGRROTNPV---PAKRAVRTQ-----SLPIW----- 280
QY 216 RVDNLNRPPTGGGYIGRNSVDMCFY-----DGYSNSSSLEIRQDNNPKSDGKFLYLR 269
QY 281 ---NKPINANTGPPRNSIAHKIYIADPKTTEANQONOHINAVEINTLQNDNTIQRK 336
QY 270 INDD--TKELAYTLSLDLACKSLPTNGT-----SLNI 300
QY 337 IDDSVSKRDFMLNKRKSKTKVSPGIATLAKKPASINI 374
RESULT 11
N33_YEAST
ID N33_YEAST STANDARD; PRT; 1157 AA.

```

```

AC p36161;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Nucleoporin NUP133 (Nuclear pore protein NUP133).
GN NUP133 OR-YKR082W OR-YKR402.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRATN-JUXUR;
RA MEDLINE=95112817; PubMed=7813444;
RA Doye V., Mepf R., Hurt E.C.;
RT "A novel nuclear pore protein Nup133p with distinct roles in poly(A)+
RL RNA transport and nuclear pore distribution.";
RL EMBO J. 13:6062-6075(1994).
RN [2]
SEQUENCE FROM N.A.
MEDLINE=94262327; PubMed=8203164;
Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
Remacha M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;
RT "The complete sequence of an 18,002 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the HBS1, MKP-U20 and PRP16 genes,
RT and six new open reading frames.";
RL Yeast 10:231-245(1994).
CC -1- FUNCTION: INVOLVED IN POLY(A)+ RNA TRANSPORT AND NUCLEAR PORE
CC DISTRIBUTION.
CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X80066; CAA56372.1; -
CC EMBL: 227116; CAA81633.1; -
CC EMBL: 228307; CAA82161.1; -
CC PIR: S38160; S38160.
CC PIR: S39123; S39123.
CC SGD: S0001790; NUP133.
CC Nucleolar protein; Transport; Transmembrane.
CC TRANSMEM 217 233
CC STIMLAR 98 420 TO NUP120 (AA 434-763).
CC SEQUENCE 1157 AA; 133319 MW; C8BDB7D709C5C08 CRC64;

Query Match 5.2%; Score 98; DB 1; Length 1157;
Best Local Similarity 22.1%; Pred. No. 8.9;
Matches 75; Conservative 43; Mismatches 127; Indels 94; Gaps 21;

QY 48 HNLNNHITAY-----SESHLYDRMTF-----LCISSNNTLNGACPTSENP 89
DB 96 YGLVNDHKVYIWNHISTORDPYITVPRSDNDDEIYAVAPRCILTFPATMDESPALNP 155
QY 90 SSSVSSE-----TNTLQFTEKRSILKRELOIK-----GKKOLLFK 126
DB 156 NDDDETGLIITIKSKAIYEDINSINNLFKISEKFS---HELELPINSSGGKCDLM- 211
QY 127 SVNC-PSGLTLNS-----AFHNCNKNAASGASLYLIPAGELKNLPFC-GIWDATLKLRL- 178
DB 212 -LNCEPGIYILSTMGRIFFTITRNSMGKPOLKL-----GKLLNKPFLGLIMSKIFNNSS 266
QY 179 -VRRRSEY--GYTINITIKLTKDGNIQIWLPOFKSDARVDLNLPTGGGTYIGNSV 235
DB 267 VVSLRNKPIIGKGRTRYIT---TNKGIFQTW-QLSA-----TNSHPT-----KLI 308
QY 236 DMCFYDGYSTNSSSL-----EIRFODNNPKSD--GKPYLRKINDDP-KEIAYVTLISLLA 286
DB 309 DVNIYEAILESLODLYPFANGTKIWDHPLQDESSQLFLSLSYDSSCNETYYILSTIIF 368

```

```

QY 287 GKSLTPNGTSLNADAALETNMRITAYTM--PEISVP 324
DB 369 DSS-----SNSFTIFSRYRLNTFMESITDTRKPKRIIFP 402

RESULT 12
CDG2_PAEMA
ID CDG2_PAEMA STANDARD; PRT; 713 AA.
AC P18835;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cyclomaltoextrin glucanotransferase precursor (EC 2.4.1.19)
DE (Cyclomaltoextrin-glycosyltransferase) (CGTase).
OS Paenibacillus macerans (Bacillus macerans).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=44252;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-37.
RA Sugimoto T., Kubota M., Sakai S.;
RT "Polypeptide possessing cyclomaltoextrin glucanotransferase
RT activity.";
RL Patent number GB2169902, 23-JUL-1986.
CC -1- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation
CC of a 1,4-alpha-D-glucosidic bond.
CC -1- COFACTOR: BINDS TWO CALCIUM IONS.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: CGTASE MAY CONSIST OF TWO PROTEIN DOMAINS: THE ONE
CC IN THE AMINO-TERMINAL SIDE CLEAVES THE ALPHA-1,4-GUCCOSIDIC BOND
CC ACTIVITIES, INCLUDING THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER
CC ALPHA-1,4-GUCCOSIDIC LINKAGE FOR CYCLIZING THE
CC MALTOOLIGOSACCHARIDE PRODUCED.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC PIR: S26589; ALBSXR.
CC HSSP: PA3379; 1CDG.
CC InterPro: IPR000461; Alpha_amylase.
CC InterPro: IPR002044; CBD_4.
CC InterPro: IPR002909; IPT_TIG.
CC Pfam: PF00128; alpha-amylase; 1.
CC Pfam: PF00686; CBM_20; 1.
CC Pfam: PF01833; TIG; 1.
CC Pfam: PF02806; alpha-amylase-C; 1.
CC Pfam: PD001568; CBD_4; 1.
CC Transfaser: Glycosyltransferase; Calcium; signal.
CC SIGNAL 1 27
CC CHAIN 28 713 CYCLOMALTODEXTRIN GLUCANOTRANSFERASE.
CC DOMAIN 28 165 A1.
CC FT DOMAIN 166 229 B.
CC FT DOMAIN 230 434 A2.
CC FT DOMAIN 435 522 C.
CC FT DOMAIN 523 609 D.
CC FT DOMAIN 610 713 E.
CC FT ACT_SITE 256 256 BY SIMILARITY.
CC FT ACT_SITE 285 285 BY SIMILARITY.
CC FT ACT_SITE 356 356 BY SIMILARITY.
CC SEQUENCE 713 AA; 76857 MW; 5A287BCC4AAFE835 CRC64;

Query Match 5.1%; Score 95.5; DB 1; Length 713;
Best Local Similarity 20.1%; Pred. No. 7.4;
Matches 65; Conservative 43; Mismatches 131; Indels 85; Gaps 12;

QY 68 MTFCLSSNHTLNGACPTSENPSSSSVSGETNITITLQFTEKRSILKRELOIKGYKOLLFKS 127
DB 8 LTVSVMSSVGIALGAALPVWASP--DTSVNNKLNFS---TDFVYQIVDRFVDG-----NS 57
QY 128 VNCPSEGLTLNSAFHNCNKNAASGASLYLIPAGELKNLPFGGIWDATLKLVRVRRSEY 187
DB 58 ANNPVGAAPSSDSHN-----LKLX-----FGGMQGS-----INNKINDGY 92
QY 188 GYTYINITIKLTKDGNIQIWLPOFKSDARVDLNLPTGGGTYIG----- 231

```



```

Db 93 -----LTGMDITALLISQPEVNTTAVIINSGVANNATVHGWPDPFKTNAAGPSF 142
Qy 232 -----RNSVDMCFYDDIGYSTSSSLSEIRFQONNRKSDCKFLRKINDTKEIA 278
Db 143 TDFSNLIAAAHSHNIKVYMDPAHPNTNTAASSTDPSEFANGALYNNGTLLGRYSNDTALF 202
Qy 279 YTLISLLAGKSLFTPE-NGTSLINADIASALESLENNMTITVTPMEISVPVLCWPGRQLDPAK 337
Db 203 HH-----NCGTPEFTSESTIYKNIYDLADIINONNNNTIDSYLKEISIDMLNLGVGDGIRPDV 258
Qy 338 VENPEAGQ--YMGNT---NVTFT 355
Db 259 KHPGQWQKSYSSISYSSANPVFT 282

RESULT 13
RF1M_YEAST STANDARD; PRT; 413 AA.
ID RF1M_YEAST
P30775;
01-JUL-1993 (Rel. 26, Created)
01-JUL-1993 (Rel. 26, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
De Peptide chain release factor 1, mitochondrial precursor (MRF-1).
MRF1 OR YGL143C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
ON NCBI_TaxId=4932.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93117110; Pubmed=1475194.
RA Pel H.J., Mat M.J., Rep M., Griwell L.A.;
RT "The yeast nuclear gene MRF1 encodes a mitochondrial peptide chain
RT release factor and cures several mitochondrial RNA splicing
RT defects.";
RL Nucleic Acids Res. 20:6339-6346(1992).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=S288C / FY1769;
RX MEDLINE=97197983; Pubmed=9046099;
RA Voelt M., Deifour E., Verhasselt P., Riles L., Robben J., Volckaert G.;
RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm
RT of chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL3A,
RT TPI1, MRF1 genes and six new open reading frames.";
RL Yeast 13:117-182(1997).
-1 FUNCTION: MITOCHONDRIAL PEPTIDE CHAIN RELEASE FACTOR THAT DIRECTS
THE TERMINATION OF TRANSLATION IN RESPONSE TO THE PEPTIDE CHAIN
TERMINATION COONS UAA AND UAG.
-1 SUBCELLULAR LOCATION: Mitochondrial.
-1 SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE
FACTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/1sb.ch).
CC or send an email to license@1sb.ch.
CC -----
Cc EMBL; X60381; CAA42932.1; -
Cc EMBL; X99960; CAA68219.1; -
Cc EMBL; Z72655; CAA96855.1; -
Cc PIR; S28602; S28602.
Cc SGD; S000311; MRF1.
Cc InterPro: IPR005139; PCRF.
Cc InterPro: IPR000352; Pep_rel_factor_I.
Cc Pfam: PF004472; RF-1; 1.
Cc Pfam: PF004462; PCRF; 1.
Cc PROSITE; PS00745; RF_PROK_I; 1.
Cc Protein biosynthesis; Mitochondrion; Transit peptide.
Cc TRANST 1 MITOCHONDRION (POTENTIAL).

```

FT	CHAIN		?	413		PEPTIDE CHAIN RELEASE FACTOR 1.
SQ	SEQUENCE	413 AA;	46770 MW;	AAD4829748C7604 CRC64;		
	Query Match		5.0%;	Score 95;	DB 1;	Length 413;
	Best Local Similarity		21.8%;	Pred. No. 4;		
	Matches 67;	Conservative 47;	Mismatches 121;	Indels 72;	Gaps	
OY	SSSVSGSTN-IILQETKRSLLI-----KRELQIKGYKQLLEFKSYNCPSGLTNSAHFN	142				
Dd	: :	11:				
Dd	30 TSTTTNKSNGSIPOTYTELSPFLVKAQEYEVELMD-----LDKDLSC-----GHFD	78				
OY	143 CNK-----NAASGASLYLYIPAGELKNLPFGGIMDATLKLRVKRRSE---TYGYTYI	192				
Dd	: :	11:				
OY	193 NITLIKLDKGNIQWMLQEFKSDARVDLNLRPTGG---TYIGNSVDMCFYCYGTNSSS	249				
Dd	139 RLNVKLDP-----PHPFADKPSSLLELRPGVGIEAMIFTONLDM--YIGYA-NRYK	187				
OY	250 LEIFFOONNPKSDKEFYLRKINDPYEKIAVTLSLLLAGK-----SLAPTNSTNLADI	303				
Dd	188 MKYATIKSKNEHSESSGIIDALISLEEGSGDRLRFEGVHRVQRIRISTETKG-----	239				
OY	304 ASLETNNRRTAVTMPEI-----SVPLVCPCGRLOLDAKVENPERGQTGMN---I	350				
Dd	240 ---RTHNSTAAVVLPQIGDESAKSIDAYERTFRKPGEIRVDIMRASCKGQHVNTTDSAV	296				
OY	351 NVETFPS 357					
Dd	: :					
Dd	297 RLTHIPS 303					
	RESULT 14					
MAPA	BACSU					
ID	MAPA_BACSU	STANDARD;		PRT; 2334 AA.		
AC	007833;					
DT	01-OCT-1994 (Rel. 30, Created)					
DT	01-OCT-1994 (Rel. 30, Last sequence update)					
DT	15-JUN-2002 (Rel. 41, Last annotation update)					
DE	Wall-associated protein precursor.					
GN	MAPA OR NI7G.					
OS	Bacillus subtilis.					
OC	Bacteria, Firmicutes; Bacillales; Bacillaceae; Bacillus.					
OK	NCB1_taxid=1423;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-168;					
RC	MEDLINE-93302506; PubMed-8316082;					
RA	Foster S.J.;					
RT	"Molecular analysis of three major wall-associated proteins of					
RT	Bacillus subtilis 168: evidence for processing of the product of a					
RT	gene encoding a 258 kDa precursor two-domain ligand-binding					
RT	protein".;					
RL	Mol. Microbiol. 8:299-310(1993).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-168 / BGSCIAL;					
RC	MEDLINE-97124196; PubMed-8969509;					
RA	Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,					
RA	Miya Y., Fujita Y.;					
RT	"Sequencing of a 65 kb region of the Bacillus subtilis genome					
RT	containing the ltc and cel loci, and creation of a 177 kb contig					
RT	covering the gnt-saxc region.";					
RL	Microbiology 141:3113-3123(1996).					
CC	-1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM.					

CC	MOTILITY, SECRETION OR DIFFERENTIATION.
CC	-1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED INTO THE MEDIUM.
CC	-1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED MOTIF REPEATED 31 TIMES.
CC	-1- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME SIMILARITY TO THE REPEAT IN E. COLI RHS GROUP OF PROTEINS (RHS-A-D).
CC	-----
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; L05634; AAA22883.1; -
DR	EMBL; D31856; BAA06556.1; -
DR	EMBL; D29885; BAA06260.1; -
DR	EMBL; D83026; BAA11683.1; -
DR	EMBL; Z99124; CAB15959.1; -
DR	PIR; S32920; S32920.
DR	Subtilist; BG10797; wapa.
DR	Interpro: IPR003305; CBM_Cenc.
DR	Pfam; PF02018; CBM_4_9; 1.
KW	Cell wall; Repeat; Signal; Complete proteome.
FT	SIGNAL 1 28 OR 32 (POTENTIAL).
FT	FT CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.
FT	FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
FT	REPEAT 504 605 1-1.
FT	REPEAT 636 735 1-2.
FT	REPEAT 769 869 1-3.
FT	DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
FT	REPEAT 1021 1040 2-1.
FT	REPEAT 1042 1061 2-2.
FT	REPEAT 1063 1082 2-3.
FT	REPEAT 1083 1102 2-4.
FT	REPEAT 1109 1128 2-5.
FT	REPEAT 1129 1148 2-6.
FT	REPEAT 1150 1169 2-7.
FT	REPEAT 1174 1193 2-8.
FT	REPEAT 1199 1218 2-9.
FT	REPEAT 1219 1238 2-10.
FT	REPEAT 1249 1268 2-11.
FT	REPEAT 1267 1286 2-12.
FT	REPEAT 1286 1305 2-13.
FT	REPEAT 1305 1324 2-14.
FT	REPEAT 1324 1343 2-15.
FT	REPEAT 1343 1362 2-16.
FT	REPEAT 1362 1381 2-17.
FT	REPEAT 1381 1400 2-18.
FT	REPEAT 1400 1419 2-19.
FT	REPEAT 1419 1438 2-20.
FT	REPEAT 1438 1457 2-21.
FT	REPEAT 1457 1476 2-22.
FT	REPEAT 1476 1495 2-23.
FT	REPEAT 1495 1514 2-24 (APPROXIMATE).
FT	REPEAT 1514 1533 2-25.
FT	REPEAT 1533 1552 2-26.
FT	REPEAT 1552 1571 2-27.
FT	REPEAT 1571 1590 2-28.
FT	REPEAT 1590 1609 2-29.
FT	REPEAT 1609 1628 2-30.
FT	REPEAT 1628 1647 2-31.
FT	REPEAT 1647 1666 2-32.
FT	REPEAT 1666 1685 2-33.
FT	REPEAT 1685 1704 2-34.
FT	REPEAT 1704 1723 2-35.
FT	REPEAT 1723 1742 2-36.
FT	REPEAT 1742 1761 2-37.
FT	REPEAT 1761 1780 2-38.
FT	REPEAT 1780 1799 2-39.
FT	REPEAT 1799 1818 2-40.
FT	REPEAT 1818 1837 2-41.
FT	REPEAT 1837 1856 2-42.
FT	REPEAT 1856 1875 2-43.
FT	REPEAT 1875 1894 2-44.
FT	REPEAT 1894 1913 2-45.
FT	REPEAT 1913 1932 2-46.
FT	REPEAT 1932 1951 2-47.
FT	REPEAT 1951 1970 2-48.
FT	REPEAT 1970 1989 2-49.
FT	REPEAT 1989 2008 2-50.
FT	REPEAT 2008 2027 2-51.
FT	REPEAT 2027 2046 2-52.
FT	REPEAT 2046 2065 2-53.
FT	REPEAT 2065 2084 2-54.
FT	REPEAT 2084 2103 2-55.
FT	REPEAT 2103 2122 2-56.
FT	REPEAT 2122 2141 2-57.
FT	REPEAT 2141 2160 2-58.
FT	REPEAT 2160 2179 2-59.
FT	REPEAT 2179 2198 2-60.
FT	REPEAT 2198 2217 2-61.
FT	REPEAT 2217 2236 2-62.
FT	REPEAT 2236 2255 2-63.
FT	REPEAT 2255 2274 2-64.
FT	REPEAT 2274 2293 2-65.
FT	REPEAT 2293 2312 2-66.
FT	REPEAT 2312 2331 2-67.
FT	REPEAT 2331 2350 2-68.
FT	REPEAT 2350 2369 2-69.
FT	REPEAT 2369 2388 2-70.
FT	REPEAT 2388 2407 2-71.
FT	REPEAT 2407 2426 2-72.
FT	REPEAT 2426 2445 2-73.
FT	REPEAT 2445 2464 2-74.
FT	REPEAT 2464 2483 2-75.
FT	REPEAT 2483 2502 2-76.
FT	REPEAT 2502 2521 2-77.
FT	REPEAT 2521 2540 2-78.
FT	REPEAT 2540 2559 2-79.
FT	REPEAT 2559 2578 2-80.
FT	REPEAT 2578 2597 2-81.
FT	REPEAT 2597 2616 2-82.
FT	REPEAT 2616 2635 2-83.
FT	REPEAT 2635 2654 2-84.
FT	REPEAT 2654 2673 2-85.
FT	REPEAT 2673 2692 2-86.
FT	REPEAT 2692 2711 2-87.
FT	REPEAT 2711 2730 2-88.
FT	REPEAT 2730 2749 2-89.
FT	REPEAT 2749 2768 2-90.
FT	REPEAT 2768 2787 2-91.
FT	REPEAT 2787 2806 2-92.
FT	REPEAT 2806 2825 2-93.
FT	REPEAT 2825 2844 2-94.
FT	REPEAT 2844 2863 2-95.
FT	REPEAT 2863 2882 2-96.
FT	REPEAT 2882 2901 2-97.
FT	REPEAT 2901

Oy	29	DESIGNINGPRNRNNSPPKHNLINN-----HITR-----YSSEHTLY---- <td>74</td>	74
Dd	260	DSKIDELSGEVERSD-KVSYIKLENKEEGLHLHDADEWMLKDPEVVYVSIDPSTSLSVS	318
Oy	75	SHNTLNGACPTSENSSSSVSGETNITLOFTEKRSLIKRELOIKGYKOLLFRSVNCPSGL	134
Dd	319	SDFEVMASVPPT-NYSASSOKWDAN-----LKAY-----	346
Oy	135	TLNASHFNCKNNMAAGBALYLILYPAGELKNLPFGIMATIKLRKKRYSETGYTTINI	194
Dd	347	VLTGTY- ----DKTGTGTVAFMFFNNLNKLPIQNMVTVAATLKYTAHSY----YGf-----	393
Oy	195	TIKLDKGNIOIWLPOFS--DARVDNLPTPGGGTFIGRNSVDMCYDYg-----S	244
Dd	394	--KAQG-----LMIDTVASNVDNAVMTNTRPKASN--IGRADVHGKGOMASYDTVAAVKS	444
Oy	245	TNSSLEIRFPQDNPNKSPGCKFYLRKI-----NDTRELAIYTLSSLLAGSLPPT-----	293
Dd	445	WNSSGANYGFK-LHTNGNGEKWEYKRLISSANSANKPYIEVYTTP-----KGNPTIKAY	498
Oy	294	-NGTSLNTADAAASLETNNMRI	313
Dd	499	HNGDSTGYFD-----ISMKV	514
 RESULT 15 N100_YEAST STANDARD; PRT; 959 AA.			
ID	N100_YEAST	STANDARD;	PRT; 959 AA.
Ac	002629;		
Dt	01-JUL-1993	(Rel. 26, Created)	
Dt	01-JUL-1993	(Rel. 26, Last sequence update)	
Dt	01-OCT-1996	(Rel. 34, Last annotation update)	
De	Nucleoporin NUP100/NSPI100	(Nuclear pore protein NUP100/NSPI100).	
Gn	NUP100 OR NSP100 OR YKL068W OR YKL336.		
Os	Saccharomyces cerevisiae (Baker's yeast).		
Oc	Eukaryota, Fungi, Ascomycota; Saccharomycotina; Saccharomycetes;		
Oc	Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.		
Ox	NCBI_Taxid=4932;		
Rn	[1]		
Rp	SEQUENCE FROM N.A.		
Rx	MEDLINE=93054906; PubMed=1385442;		
Rt	Wente S.R., Rout M.P., Blobel G.;		
Rl	"A new family of yeast nuclear pore complex proteins.";		
Rl	J. Cell Biol. 119:705-723(1992).		
Rn	[2]		
Rp	SEQUENCE FROM N.A.		
Rc	STRAIN=S288C;		
Rx	MEDLINE=94378724; PubMed=8091863;		
Ra	Rasmussen S.W.;		
Rt	"Sequence of a 20.7 kb region of yeast chromosome XI includes the		
Rt	NUP100 gene, an open reading frame (ORF) possibly representing a		
Rt	nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in		
Rt	addition to seven ORFs with weak or no significant similarity to		
Rt	known proteins."		
Rl	Yeast 10:569-S74(1994).		
Cc	-I- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.		
Cc	NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF		
Cc	THE PROTEIN DURING NUCLEOCYTOPLASMIC TRANSPOrt.		
Cc	-I- SUBCELLULAR LOCATION: Nuclear pore complex.		
Cc	-I- DOMAIN: CONTAINS G-L-F-G REPEATS.		
Cc	-I- SIMILARITY: BELONGS TO THE GEG FAMILY OF NUCLEOPORINS.		
Cc	This SWISS-PROT entry is copyright. It is produced through a collaboration		
Cc	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
Cc	the European Bioinformatics Institute. There are no restrictions on its		
Cc	use by non-profit institutions as long as its content is in no way		
Cc	modified and this statement is not removed. Usage by and for commercial		
Cc	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
Cc	or send an email to license@sib-sib.ch).		
Dr	EMBL; Z15035; CAA78753.1; ;		
Dr	EMBL; X75780; CAA53406.1; ;		
Dr	EMBL; Z28068; CAA81905.1; ;		

DR PIR: B44402; B44402.
DR PIR: S39173; S39173.
DR PIR: S44518; S44518.
DR SGD: S0001551; NUP100.
DR InterPro: IPR004325; Nucleoporin_FG.
DR Pfam: PF03093; Nucleoporin_FG; 24.
KW Nuclear protein; Transport; Repeat.
FT DOMAIN 33 571
G-L-F-G
SQ SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;

Query Match 5.0%; Score 94.5; DB 1; Length 959;
Best Local Similarity 19.5%; Pred. No. 13;
Matches 80; Conservative 58; Mismatches 157; Indels 115; Gaps 17;

QY 31 SITNIFGPPDRNESS-----PKHNIL-----NNHTAYSESHLYDRMFLCLSSHNTLNG 81
DB 284 SMSGLEFGQSMNSSTOGVFGONNNQMOINGNNNSLFGKANTFSNSASGGLFGQNNQOOG 343
82 ACPTSENPSSSVSG-----EENITLOFTEKRSLIKRELQIKGYK----- 121
DB 344 SGLFGONSGTSGSSGLFGONNOKOPNTFTQSTNGIGL-FGQNNNOQOOSTGLFGAKPACT 402
QY 122 -----QLLEKSVNCPGSLTLNSAHFNCNKNAASGASLYLYIPAGELKNLPPG 168
DB 403 TGLFGGNSSTQPNSLFGTTNPTS-----NTQSQGNSLFG--GATKLTNMPG 449
QY 169 GIWDATLKLIRVRRYSEYGYTTINTITIKLTDKNIQIWLPOKFS----- 213
DB 450 GNPRTA-----NOSGSGNSLFGTKPASTTGLFGNNTASTVTPSTNGLFGNNANNSTSTNT 505
QY 214 ---DARVDLNLRLPT--GGGY-----IGRN-----SVMCPYDGYSTNSSSL--- 250
DB 506 GLFGAKPDSQSPALGGGLFGNSNSNSTIGOKPVFGCTTONTGLFGATGNSSAVGST 565
QY 251 -EIRFQDNMPKSDGKFEYLRKINDTKEIAYTLSSLAGKSLPTNGTSLINIAAASLETN 309
-DB 566 GRLFGQNNNTLNVGTQNVPPVNNNTQN-----ALLGTTAVP---SLQQApy-----TN 610
QY 310 WNRITAVTMP-EISVPVLCWPGRLQIDAKVENPEAGQYMGININVTTPSS 358
DB 611 EQLFSKISIPNSITNPVKATTSKVNADMKRNSLSAVRYLAPKPLFAPSS 660

Search completed: December 4, 2002, 17:27:58
Job time : 19 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2002, 17:26:37 : Search time 21 Seconds
(without alignments)
1652.596 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886

Sequence: 1 MNKILFIFTLFFSSVLETF.....EAGQYMGINIVTFPPSSQTL 361

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1558.5	82.6	360	2 D56617	cfaf protein precu
2	949.5	50.3	363	2 S49539	COOD protein precu
3	930.5	49.3	364	2 S57937	CotD protein precu
4	275.5	14.6	359	2 AE0541	probable fimbrial
5	120	6.4	1335	2 G90975	probable factor [i
6	120	6.4	2660	2 E85832	probable invas
7	114	6.0	650	2 S22835	alpha-agglutin
8	109	5.8	2383	2 D64962	probable membra
9	108	5.7	524	2 S55097	probable membra
10	107	5.7	2529	2 B64635	toxin-like outer m
11	107	5.7	304	2 A64904	probable fimbrial
12	107	5.7	304	2 C90892	probable adhesin
13	107	5.7	304	2 P85735	probable adhesin
14	106	5.6	671	2 A38109	autolysin - Entero
15	105.5	5.6	843	2 AC2507	hypothetical prote
16	105	5.6	598	2 AB1236	internalin protein
17	104.5	5.5	1441	2 B86807	hypothetical prote
18	103.5	5.5	691	2 B75622	hypothetical prote
19	103.5	5.5	1269	2 A90267	proteinase related
20	103.5	5.5	2399	2 H71879	toxin-like outer m
21	102	5.4	692	2 G90284	hypothetical prote
22	101	5.4	1238	2 T48965	probable exported
23	100	5.3	1752	2 T48965	hypothetical prote
24	99.5	5.3	5627	2 C83339	hypothetical prote
25	99	5.2	802	2 AH1580	phenylalaninyl-trna
26	99	5.2	856	2 B81389	probable periplasm
27	99	5.2	1849	2 C41859	Iga-specific metal
28	98.5	5.2	608	2 H64473	hypothetical prote
29	98.5	5.2	4919	2 T31105	hypothetical prote

30	98	5.2	587	2 AC1510	internalin protein
31	98	5.2	758	2 S47866	SPT21 protein - ye
32	98	5.2	1014	2 S37405	Cytotoxic necrotiz
33	98	5.2	1157	2 S38160	NBP133 protein - y
34	97.5	5.2	1797	2 F69195	cell surface glyco
35	97	5.1	454	2 T01337	hypothetical prote
36	96.5	5.1	2364	2 T40884	Cytotoxin L - Cios
37	96	5.1	992	2 T28421	probable DNA-direc
38	95.5	5.1	713	2 A1BSXR	Cyclomalodextrin
39	95.5	5.1	895	2 T02397	Mutator-like trans
40	95	5.0	413	2 S28602	translational releas
41	95	5.0	561	2 A84113	transposase (12) B
42	95	5.0	2334	2 S32920	cell wall-associat
43	95	5.0	5188	2 B85547	probable RTX fami
44	94.5	5.0	835	2 E71691	outer membrane ass
45	94.5	5.0	858	2 T47223	replication licens

ALIGNMENTS

RESULT 1	
D56617	cfaf protein precursor - Escherichia coli plasmid NTP113
C:Species: Escherichia coli	
C:Date: 05-Jan-1996	#sequence_revision 05-Jan-1996 #text_change 10-Dec-1999
C:Accession: D56617	
R:Jordi, B.J.; Willshaw, G.A.; van der Zeijst, B.A.; Gastra, W.	
DNA Seq. 2, 257-263, 1992	
A:Title: The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of	
A:Reference number: A56617; MUID:92329981; PMID:1352712	
A:Accession: D56617	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-360 <JORD>	
A:Cross-references: GB:M55661; NID:q145507; PIDN:AAC41417.1; PID:q145511	
A:Experimental source: enterotoxigenic strain, CFA/I-ST plasmid NTP113	
A>Note: sequence extracted from NCBI backbone (NCBIRN:108960, NCBIRP:108972)	
A:Genetics:	
A:Gene: cfaf	
A:Genome: plasmid	
C:Superfamily: Escherichia colonizing factor antigen cfaf	
Query Match	82.6% Score 1558.5; DB 2; Length 360;
Best Local Similarity	81.2% Pred. No. 1.3e-109;
Matches 293; Conservative 31; Mismatches 36; Indels 1; Gaps 1;	
QY 1 MNKILFIFTLFFSSVLETFPVSADKIPGDESITNFGPRDRNESSPKHNLNHTAYSE 60	
DB 1 MNKILFIFTLFFSSGFFFPVVSADANPCSENMNTNIGPHDRGSSPIYNILSYLTAYNG 60	
QY 61 SHTLYDRMTFLCLSSHNLTNGACPTSENPSSSSVSGEFNTIQTETKRSLLKRELQIKGY 120	
DB 61 SHHTLDKMSFLCLSSQNTLNGACPSDDAPGTATIDGETNITLQETKRSLLKRELQIKGY 120	
QY 121 KQLFKSVNCPGGLTNSAHFMCNKAASGASLYITPAGELKNLPFGGIMDATKLVK 180	
DB 121 KQLEKKNACPSKIALNSHFCQNEQASGATLSITPAGELKNLPFGGIMAVATKLVK 180	
QY 181 RRYSETGYTYTINIRIKTKDGNIOIWLPOFKSADAVDLNLPPTGGTYIGNSVDMCFY 240	
DB 181 RRYDTTYGTITNITVINLTKDGNIOIWLPOFKSNARVDLNLPTGGTYIGNSVDMCFY 240	
QY 241 DGYSTNSSLETRFODNNPKSDKRYLKRINDYKEIAYTSLSLAGSKSLPTNGTSANI 300	
DB 241 DGYSTNSSLETRFODNNPKSDKRYLKRINDSKELYTSLLAGSKSLPTNGQALANI 300	
QY 301 ADAASLETNNMTITAVTPEISVPLCWGRLOLAKYENPAGQYMGINIVTFPPSSQT 360	
DB 301 -NTASLETNNMTITAVTPEISVPLCWGRLOLAKYENPAGQYMGINIVTFPPSSQT 359	
QY 361 L 361	

Db 360 L 360

RESULT 2

S49539

C:CoD protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999

C:Accession: S49539

R:Prochlich, B.J.; Karakashian, A.; Melsen, L.R.; Wakefield, J.C.; Scott, J.R.

Mol. Microbiol. 12, 387-401, 1994

A:Title: Cooc and Cooc are required for assembly of CS1 pili.

A:Reference number: S49538; M01D:94344028; PMID:7915003

A:Accession: S49539

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-363 <PRO>

A:Cross-references: EMBL:X76908; NID:g488735; PIDN:CAA54230.1; PID:g488737

C:Superfamily: Escherichia colonizing factor antigen cfaB

Query Match

Best Local Similarity 50.3%; Score 949.5; DB 2; Length 363;

Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;

Db 3 KILFIPLFFSVLFFAVSADKIPGDE--STNIF-GPR-DRNESSPKHNIINHTAY 58

Db 2 KKIIFL---SIIFSAVVSAGRYPTTVGNLTKSFQAPLRDSVQSPIYINIFTHVAGY 57

Db 59 SESHTLYDRMTFLCLSHNTLNGACPTSENPSSSVS-GETNTIOTTEKRSILKRELQI 117

Db 58 SLHSLDLRIVFLCTSSSNVNGACPTI--GTSVQYGTITTLTLOTEKRSILKRNINL 114

Db 118 KGYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 174

Db 115 AGAKKPIWENQSDFSMLAMVNLKNSMCGAHGANGNLMLYIPAGINKLPGCIWEAT 174

Db 175 LKLRKRRSET---YGYTTITITKLDKGNIQIWLPOFKSDARVDLMLRPTGGTYI 230

Db 175 LILRLS-RYGEVSSHYGNTVNIITVDLTDKGNIQVWLPGFHSNPRVDLMLRPTGNYKS 233

Db 231 GRNSVMCFDGYSTNSSLSEIRPDNNPKSDCKEYLRKINDTKELAVTLISLLAGKSL 290

Db 234 GNSSLMCFDGYSTNSDSVIKFQDDNPNSESENLTKRG-GETKLPYAVSLMGEKIF 292

Db 291 TPTNGSLNIAADAASLETNNRITTAVTMPETISVPLQMPRLQDLDAKVENPEAGQYVNI 350

Db 293 YPVNGQSFINDSSVLETNNRITTAVTAMPVENVPLQMPARLLMLNDVNPADQGYSGOI 352

Db 351 NVTFPSSQTL 361

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Db 353 YITFTPSVENTL 363

Best Local Similarity 48.4%; Pred. No. 2e-62; Matches 177; Conservative 64; Mismatches 118; Indels 7; Gaps 3;

Db 1 MNKILFIPLFFSVLFFAVSADKIPGDE--STNIF-GPR-DRNESSPKHNIINHTAY 58

Db 1 MKKIVFLSWFLCSQYVGSWHNVNAGSINKTESICPIDRSAASTPAYITHEHAGY 60

Db 59 SESHTLYDRMTFLCLSHNTLNGACPTSENPSSSVSGETNTIOTTEKRSILKRELQI 118

Db 61 NKDHSLEDRMTFLCLSHNTLNGACPTSENPSSSVSGETNTIOTTEKRSILKRELQI 118

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Db 119 GYKQLLFKSVNCPG--LTLSAHFNCKNKA-ASGASLYLIPAGELKMLPFGIMDAT 175

Best Local Similarity 29.7%; Score 275.5; DB 2; Length 359;

Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;

Db 70 FLCLSSHNTLNGACPTSENPSSSVSGETNTIOTTEKRSILKRELQI 129

Db 79 WVCRRNNENEGACEETHLWVWYAFGYSKIRLRFRQISHABITL-----ILGSVR 131

Db 130 --CPGSLTNSAHFNCKNKAAS--GASLYLIPAGELKMLPFGIMDATILK-RVKRRY 183

Db 132 DACYTG-----INMNAACOMGRSLKLRIPSEELIATIPISGWKATIVDIYQMG 183

Db 184 SETYGYTITITKLD--KGNQIWLPOF-KSDARVDLMLRPTGGGYIGRNSVMCEY 240

Db 184 DPLGISTDITLNVGDHFAENAIYFPGGTATPRVDLMLHRMNSOMSGRANLDMCY 243

Db 241 DGSTNSSLSEIRFQDNNPKSDCKEYLRKINDTKELAVTLISLLAGKSLTPTNGSLN 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300

Db 244 DG-GYKARSLQMKIESNKSNGTSGFYIKSDSADT--IDYAVSNMNGRSIPVTRGVEFSL 300</

```

OY 301 ADASLENNMRITAVTMEPEISVPYLCPMGRQLDAAK---VENPAGGYMGNIINTFTPPS 357
Db 301 DWVDKAAAR-----PVYLPGRQAVRCVPELTLTTTQFPNIREKRSGBEYOGTLVTMLMG 355
OY 358 SQT 360
:|
Db 356 TQT 358

RESULT 5
G90975
probable factor [imported] - Escherichia coli (strain O157:H7, substrain R1MD 0509952)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: G90975
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasavara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hatfort, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
C:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic
:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90975
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1335 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA36198.1; PID:G13362243; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
:Gene: ECs2775

```

```

Query March 6.4% Score 120; DB 2; Length 1335;
Best Local Similarity 20.4%; Pred. No. 0.86; Mismatches 156; Indels 90; Gaps 15;
Matches 80; Conservative 66;

QY      13 SSVLEFFAVSADKIPGDESITNIFGPRDRNESSPKHNLNNHTAY---SESHLYDRMT 69
          :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db       746 SAKITLASANNVGLANENANTVASVNAVDGS--NPINDHTVFAVLSSGATSATSENNON 802
          ::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY      70 FLCLSSHNTLNGACPTSPNPSSSVSGETNITLOPTEKRSLI-----K 112
          ::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db      803 ----TAKTDVNGLA-TFDLKSSKQEDNTEVTLEGVNQTLIVSFVGDSSTPAQYDLQSK 857
          ::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY     113 RELQIKGY-KOLLFEKSYVCPSGLITLSAHFNCKNKAAASGASYLIYPAGELKNLPFGGIW 171
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      858 NEVVADGNDSDMTMTATVADAKGNLNLDVKYTFENVSAALKSQTPEVNSHD-----GIA 910
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY     172 DATLKLVRKKRRYSSETGYTTINITYIKLTDKGNIQIWLDPFKSDARVDILNRLPGGGTYIG 231
          |||::||::||::||::||::||::||::||::||::||::||::||::||:
          911 TATLL-----SLKNGDYTVTASVSSGSQAANOQYFIQDGSTAALTSLV-PSGDITV-- 960
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY     232 RNSVDMCFYDGISTFSSSLE----IRPDONPNKSGRKRYLKIINDMEIAYTISLLAG 287
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db      961 -----TNTPAPLHMATATLQDKNGNPLKKEITFSVPND---VASRETSINSNG 1003
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY     288 KSLPEPTNGTSL-----NIDAAASLETNMNRITAVTMPEISVPVIC 327
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     1004 KGMDSNCTATALASLGLTLAGTHMITARLANSVSDPTQMFTYVADKDRAVVYLQTSKAIII 1063
          :|::||::||::||::||::||::||::||::||::||::||::||::||:
QY     328 WFG--RIOLDARKVENPEAGQTMGININTVFPFS 357
          |::||::||::||::||::||::||::||::||::||::||::||::||:
Db     1064 GNGVDETTLTATVKDP-FDNVYKKNLSYVERFS 1094
          |::||::||::||::||::||::||::||::||::||::||::||::||:

RESULT 6
Eg5822
Probable Invasin Z3135 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85822
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Miller, L.; Grobecek, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca
Murrele 409, 529-533, 2001
Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
```

A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 12660 <SMD>
A:Cross-references: GB:AEO05174; NTD:g12516151; PIDN:ANG57041.1; GSPDB:GN00145; UWGP
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3135

Query Match Similarity 6.4%; Score 120; DB 2; Length 2660;
Best Local Similarity 20.4%; Pred. No.2.1; Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;

OY 13 SSVLTFPAVSADKIGDSITNIFGPRDRNESSPPHHNLINHTIYV---SESHLYDRLMT 69
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 737 SAKIKTILASNNNGVALINEMAATVSUNVADEGS---NPINDHVTFFAVLSSGATSFPNNON 793
OY 70 FICLISSHNTLNGACPTSENPSSSSVSGETNITLTQTEKRSLI-----K 112
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 794 ---TAKTDVNGLA-TFDLKSSKQEDNTVEYLLENGVKQTLLVSFVGSDSTQAQDLQSK 848
OY 113 RELQIKGV-KQLLFPSVNCOPSGTLNSAHFNCKNKMAAGSALYLTIIPAGELKNLPFGIM 171
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 849 NEVVADGDNSATMTATVRDAKANLNLDYKVYFPNVNSAAKLSQTEVNSHD-----GIA 901
OY 172 DATFLRKVRKRYSEFYGTITNTIKTLIDKGNIQIWLPOFKSDAHDVNLIRPFGCGTYIG 231
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 902 TATLV-----SLANGDYTVASVSSSQANQVIPIGDOSTALTILSY-PSGDITY-- 951
OY 232 RNSVDMCFDGYSTNSSLE---IRPDNNPKSPGCKFLRKINDTGEIAYTSSLLAG 287
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 952 -----TFTAPLMHTATLQDKRNGPLKDKREITFSVPND---VASRFISINSNG 994
OY 288 KSLFTPNGSL-----NIADALETMMNRITFAVTMEISVPVIC 327
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 995 KGMTDSNGIALSLTGTLTAGTHMITARLANSVSTQPRTPEYADVADKRVAVLQTSKAETI 1054
OY 328 WPG--RLQIDAKVENPEAGQVMGINIVTFTPS 357
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1055 GNGVDETTLTATVTRDP-FDNVVKNLISVFERTS 1085

RESULT 7
S22835
alpha-agglutinin - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: 22k glycoprotein; protein J1418; protein YJR004C
C:Species: Saccharomyces cerevisiae
C:date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 29-Oct-1999
C:Accession: S22835; S51229; A32822; S55192; S57019
R:Hausser, K.; Tanner, W.
FEBS Lett. 255, 290-294, 1989
A:title: Purification of the inducible alpha-agglutinin of *S. cerevisiae* and molecular
A:Reference number: S22835; MUID:90005993; PMID:2676603
A:Accession: S22835
A:Molecule type: DNA
A:Residues: 1-650 <HA>
A:Cross-references: EMBL:X16861; NID:g3352; PIDN:CAH34752.1; PID:g3353
A:Accession: S51229
A:Molecule type: Protein
A:Residues: 20-24 <HA>
R:Lipke, P.N.; Wojciechowicz, D.; Kurjan, J.
Mol. Cell. Biol. 9, 3155-3165, 1989
A:title: AG-alpha-1 is the structural gene for the *Saccharomyces cerevisiae* alpha-agglutinin
A:Reference number: A32822; MUID:90014768; PMID:2677666
A:Accession: A32822
A:Molecule type: DNA
A:Residues: 1-448,'E',450-555,'E',557-580,'L',583-650 <LRP>
A:Cross-references: GB:M28164; NID:g171041; PIDN:AAA34417.1; PID:g171044
R:de Haan, M.; Smits, P.H.M.; Grivell, L.A.,
submitted to the EMBL Data Library, May 1995
A:Reference number: S55192
A:Accession: S55192

```

A:Molecule type: DNA
A:Residues: 1-650 <DEH>
A:Cross-references: EMBL:X87611; NID:9854567; PIDD:CAA60926.1; PID:9854577
R:Jde Haan, M.; Grivell, L.A.; Smits, P.H.M.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56771
A:Accession: S57019
A:Molecule type: DNA
A:Residues: 1-650 <ZAG>
A:Cross-references: EMBL:Z49504; NID:91015625; PIDD:CAA89526.1; PID:91015626; MIPS:YJRO0C
C:Gene(s):
A:Gene: SGD:SAG1; AGAL1; AGALPHA1
A:Cross-references: SGD:S0003764; MIPS:YJRO04C
A:Map position: 10R
C:Keywords: glycoprotein

Query Match      6.08; Score 114; DB 2; Length 650;
Best Local Similarity 20.98; Pred. No. 0.95;
Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

4 ILFFTLFFSSVL---FTFA-----VSADKIPG-----DESTINIFGPRDRE--- 43
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
8 IIMFSLALSAIINIDTFPSNLETFPLANKQPOGWATPFDLSADASSINEGDEFTL 67
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
44 SSP---KHNINNHITA-----YSESHLYDRMTFLC-----LSNHTLN 80
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
68 SMPHYRIKILNNSQTATISLADGEAFKCYVSOQAAYLYENTTFCTAQNIDLSYNTID 127
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
81 GACPTSENPSSSVSGENTITLQFTEKRSILKRELQIKGYKQLLEFKSYNC--PSGLTNSA 139
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
128 GSITFSLNFSDSGGSYYEEL-----ENAKFKSGPMLVYKLGOMQSYVNFDPAPFTENVF 182
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
140 HFNCNKNAASGASLYL--YIPAGELKNLPFGGIMPATLKLRYKRRYSEYGIYTNITIK 197
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
183 HSGRSTGYGSPESYHLGMVCPNGYF---LGS-----TEKI-----DYDSSNNNVDL- 225
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
198 LTKGNLQI-----WLPQFKSDARVDNLNRPFGGTYIGRN---SYDMKCFYDG--- 242
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
226 --DCSSVOYVYSSNDPNDMWPFQSYNDTNADV---TCFSSNLMTLDEKLYDGEML 275
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
243 -----YSTNSSLE--IRFQDNN--PKSDGKF 265
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
276 WVNALQSLPAWNTIDHALFQYTCLOTLIANTYATQFSTREFIYQGRNLGTASAKSS 335
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
QY 266 YLRKINDTKI---AYLS-----LLAGKSLTPINGTSLNADAA 304
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 336 FISTTTTDLASINSAVSTGSIIVTEGNRTTSEVISHVYVTTSTKSLPATTSLLTIAQTS 395
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
305 SLETNMN-----RITAVTPEISVPVLCMPGRL 332
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 396 IYSDSNITVGTCDIHTTSEVSDVETISRETASTVAAPTSTGTGWGMAM 444
      |||:| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |

RESULT 8
D64962
probable membrane protein b1978 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: D64962
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: D64962
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2363 <BLAT>
A:Cross-references: GB:AE000289; GB:000096; NID:91788285; PIDD:AACT5042.1; PID:91788288;
A:Experimental source: strain K-12, substrain M61655
C:Keywords: nucleotide binding; P-loop; transmembrane protein
F:54-70/Domain: transmembrane #status predicted (TM)
F:1564-1571/Region: nucleotide-binding motif A (P-loop)

```

[illegible]


```

0Y 210 OERKSDARVDL-----LRPT-----GGGYTIGNSVDMCFYOSYINSSSLE----- 251
Db 472 -FISQONIDISGASALLPPTSTTSTFPGHGGGNIDINTLNTLIISNGGISSTLSAGKA 530
0Y 252 -----IRFODNNPK--SDGKFYLRKINDPTKEIAYTSLSLACKS----- 289
Db 531 GNISINSNSINVTGNTINSNSPFISSNPL--VDPNQKLLYRQPPPLLIGQAINFL 568
0Y 290 -----LTPNGTSLN-----IADAASLETMMNRITAYTMPEISVPVLCP--GRJQLDAYKE 339
Db 569 NTDIINISNGGLIARNEGVADNACINIRISANTININOGGEVNAATTTFIGEGGNITLSNR-- 646
0Y 340 NPEAGOVYGNINVTET 355
Db 647 -----NLEPLNNSRTYAT 658

```

Search completed: December 4, 2002, 17:29:07
Job time : 25 secs

This Page Blank (uspto)

Db 181 RRYSEYGYTYITITIKLADKNGIQIWLPOFKSDARVDLNLPTGGGTYIGRNSVDMCFY 240
QY 241 DGYSTNSSSLEIRFODNNPKSDGKFLRKINDPTKEIATYLSLLAGSKLPTNGTSLNI 300
Db 241 DGYSTNSSSLEIRFODNNPKSDGKFLRKINDPTKEIATYLSLLAGSKLPTNGTSLNI 300
QY 301 ADAASLETNNRITAYATMPETISVPLVLCWPGRLQDLDAKVENPEAGQYMGINIVTFPSSQT 360
Db 301 ADAASLETNNRITAYATMPETISVPLVLCWPGRLQDLDAKVENPEAGQYMGINIVTFPSSQT 360
QY 361 L 361
Db 361 L 361

RESULT 2

US-09-839-894-28
; Sequence 28, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Alboum, Zeey
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE OF INVENTION: CSA OPERON
; FILE REFERENCE: USFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 361
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
US-09-839-894-28

Query Match 98.8%; Score 1862.5; DB 9; Length 361;
Best Local Similarity 99.4%; Pred. No. 3e-161;
Matches 359; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 MNKILFIFLTFSSVLTFAVSADKIPGDESITINIFGPRDRNSSPKHNLNNHITAYSE 60
2 MNKILFIFLTFSSVLTFAVSADKIPGDESITINIFGPRDRNSSPKHNLNNHITAYSE 61
Db 61 SHLYDRMFLCLSSNHTLNGACPTSENPSSSVSGEITITLQFTEKRSLIKRELQIKGY 120
62 SHLYDRMFLCLSSNHTLNGACPTSENPSSSVSGEITITLQFTEKRSLIKRELQIKGY 121
QY 121 KOLLFESVNCPSGLTILNSAFHNCNKNAASGASLYLYIPAGELKMLPFGGIMDATLKLRYK 180
Db 122 KOLLFESVNCPSGLTILNSAFHNCNKNAASGASLYLYIPAGELKMLPFGGIMDATLKLRYK 181
QY 181 RRYSEYGYTYITITIKLADKNGIQIWLPOFKSDARVDLNLPTGGGTYIGRNSVDMCFY 240
Db 182 RRYSEYGYTYITITIKLADKNGIQIWLPOFKSDARVDLNLPTGGGTYIGRNSVDMCFY 241
QY 241 DGYSTNSSSLEIRFODNNPKSDGKFLRKINDPTKEIATYLSLLAGSKLPTNGTSLNI 300
Db 242 DGYSTNSSSLEIRFODNNPKSDGKFLRKINDPTKEIATYLSLLAGSKLPTNGTSLNI 300
QY 301 ADAASLETNNRITAYATMPETISVPLVLCWPGRLQDLDAKVENPEAGQYMGINIVTFPSSQT 360
Db 301 ADAASLETNNRITAYATMPETISVPLVLCWPGRLQDLDAKVENPEAGQYMGINIVTFPSSQT 360
QY 361 L 361
Db 361 L 361

RESULT 3

US-09-839-894-29
; Sequence 29, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Alboum, Zeey
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE OF INVENTION: CSA OPERON
; FILE REFERENCE: USFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 359
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
US-09-839-894-29

Query Match 81.8%; Score 1542; DB 9; Length 359;
Best Local Similarity 80.9%; Pred. No. 3.e-132;
Matches 292; Conservative 32; Mismatches 35; Indels 2; Gaps 2;

QY 1 MNKILFIFLTFSSVLTFAVSADKIPGDESITINIFGPRDRNSSPKHNLNNHITAYSE 60
Db 1 MNKILFIFLTFSSVLTFAVSADKIPGDESITINIFGPRDRNSSPKHNLNNHITAYSE 60
QY 61 SHLYDRMFLCLSSNHTLNGACPTSENPSSSVSGEITITLQFTEKRSLIKRELQIKGY 120
Db 61 SHLYDRMFLCLSSNHTLNGACPTSENPSSSVSGEITITLQFTEKRSLIKRELQIKGY 120
QY 121 KOLLFESVNCPSGLTILNSAFHNCNKNAASGASLYLYIPAGELKMLPFGGIMDATLKLRYK 180
Db 121 KOLLFESVNCPSGLTILNSAFHNCNKNAASGASLYLYIPAGELKMLPFGGIMDATLKLRYK 180
QY 181 RRYSEYGYTYITITIKLADKNGIQIWLPOFKSDARVDLNLPTGGGTYIGRNSVDMCFY 240
Db 181 RRYSEYGYTYITITIKLADKNGIQIWLPOFKSDARVDLNLPTGGGTYIGRNSVDMCFY 240
QY 241 DGYSTNSSSLEIRFODNNPKSDGKFLRKINDPTKEIATYLSLLAGSKLPTNGTSLNI 300
Db 241 DGYSTNSSSLEIRFODNNPKSDGKFLRKINDPTKEIATYLSLLAGSKLPTNGTSLNI 300
QY 301 ADAASLETNNRITAYATMPETISVPLVLCWPGRLQDLDAKVENPEAGQYMGINIVTFPSSQT 360
Db 300 -NTASLETNNRITAYATMPETISVPLVLCWPGRLQDLDAKVENPEAGQYMGINIVTFPSSQT 358
QY 361 L 361
Db 359 L 359

RESULT 4

US-09-839-894-31
; Sequence 31, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Alboum, Zeey
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: USFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894

```

:
: CURRENT FILING DATE: 2001-04-20
: PRIOR APPLICATION NUMBER: 60/198,626
: PRIOR FILING DATE: 2000-04-20
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 31
: LENGTH: 362
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: ETEC Protein Homology Sequence
: OS-09-839-894-31

```

Query Match	49.78;	Score 937;	DB 9;	Length 362;
Best Local Similarity	53.18;	Pred. No. 2.3e-77;		
Matches 197;	Conservative 54;	Mismatches 98;	Indels 22;	Gaps 12

Qy	3	KILFIFLFFSSVETFAVSADKIPGDE--SITNIF-GPR-DRNESSPKNINLHTAY	58
Db	2	KKIFIFL-----SIFSAVSAAGRPETTYGMLTKSPQARLDRSVGPYINIFTNHAGY	57
Qy	59	SESHLDRMTFLCLSSHTNLNGACPLSENPSSSVS-GETNITLQETEKRSKLRELQI	117
Db	58	SLSHSLYDRIFLCTSSSNPNYGACPTI---GTSGVAYGTITITLQETEKRSKLKRNINI	114
Qy	118	KGYQLLEFKSVNCSG--LTLNSAHFQCNKNA-ASGASLYLYIPAGLKLNLPGGIMDAT	174
Db	115	AGNKKPIWENOSCFPSNIMVLNSWSGCAAGNANGTILNLYIPAGIEINKLPGGIEWAT	174
Qy	175	LKLNRKRRYSRT---YGTVINTIKLTQKGNQIWLPOFKSDARVDLNLPTGGGTYL	230
Db	175	LILKLS-RYGGVSSHTNGNTVNTITVLDLTQKGNQIWLPGFHSNPRVDLNLPTGNKYS	233
Qy	231	GRNSVDMCFYDGYSTNSSLEIRQDNNPKSDGKFYLRKINDDTKEIAYTLSTLLAGKSL	290
Db	234	GSNSLDMCFYDGYSTNSDSMWIKRQDDNPNFSSSEYNLYKIG-GTEKLPYAVS-LIGKIF	291
Qy	291	TPPLNGTSLNTADAASLETNNRITATVAMPETISVPVLCWPGRLQDAKVENPEAGQIWGNT	350
Db	292	YPVNGGSEFTINDSSVLETFNNNRVATVAMPEYVNVLCWPARLLILADYVNPADAGVSGQI	351
Qy	351	NWTFPSSQTL	361
Db	352	YTFPSSVENT	362

RESULT 5
US-09-839-894-30

```

? GENERAL INFORMATION:
? APPLICANT: Alboum, Zeev
? APPLICANT: Barry, Ellen M.
? APPLICANT: Levine, Myron M.
? APPLICANT: University of Maryland
? TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
? TITLE OF INVENTION: CSA OEPRON
? FILE REFERENCE: UOFMD.006A
? CURRENT APPLICATION NUMBER: US/09/839,894
? PRIORITY FILING DATE: 2001-04-20
? PRIOR APPLICATION NUMBER: 60/198,626
? PRIORITY FILING DATE: 2000-04-20
? NUMBER OF SEQ ID NOS: 40
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 30
? LENGTH: 364
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: ETEC Protein Homology Sequence
? US-09-839-894--30

```

Query match	48.68;	Score 916.5;	DB 9;	Length 364;
-------------	--------	--------------	-------	-------------

Best Local Similarity 48.1%; Pred. No. 1.7e-75;
Matches 176; Conservative 63; Mismatches 120; Indels 7; Gaps 3;

```

QY      1 MNKLIETFEFFSVLETFVAVSADKIPGDEITINIFGPORN--ESSPKNIILNHTTAY 58
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1 MKKIIVPLSMFLCQVYGQSHHTVEGSIKTFSTIPIDRSAASTPAHYIFHEBVAGY 60
QY      59 SESHTLYDRMTFLCLSSHTNLNGACPTSENPSSSVSGETNITLQFIEKRSILIKRELOIK 118
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      61 NKDHSFLFRMTFLCMSGSPADSKGACCPGEMSKSS--GGEINIKILIEFKSSLARKTLNTLK 118
QY      119 GYKOLIEKRSVNC--PFGSLTINSAHFEMCKNKAASGASLYLIPAGELKNLPFGGIWDATL 175
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      119 GYKRFLEYESDCIHYVDKMNINSTHTVCVGSFTFGVFTFLYIPGGEIDGLTGGIMRATL 178
QY      176 KLRKRRSEIETGYTINITIKLKDCKNIQIOWIQFQKSDARVDNLNLPFGGTYIGNSV 235
      179 ELRRKRHDYHGHYKVNIIVYDLTDKNGIQOWTPKFSHPDRIIDLNPBEGGKXGSSNVL 238
QY      236 DMCFYDGYSTNSSSLLEIRFODNNRKSQGEFLRKINDDTKEIATYLSLLAGKSLPTNG 295
      239 EMCLYDGYSTHSSQSIEMRFQODSQGTNNENYLLITGEPLKRLPKXLSLLGGREFYNNG 298
Db      296 TSLNIADAASLETWNRIITAVTAPDEISVYLCWGRQLQDLAKVENPAGQMGINIVTFT 355
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      299 KAFPIINDTSSLEFIWMNRISVSLPOISIVPLCWPRANLTFMSELNNPBEAGEGSLINTFT 358
Db      356 PSSQTL 361
      359 PSSSSL 364

```

RESULT 6
 US-09-839-894-32
 Sequence 32, Application US/09839894
 Patent No. US2002017668A1
 GENERAL INFORMATION:
 APPLICANT: Alboum, zeev
 APPLICANT: Barry, Elieen M.
 APPLICANT: Levine, Myron M.
 APPLICANT: University of Maryland
 TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
 TITLE OF INVENTION: CSA OPERON
 FILE REFERENCE: USFMD 006A
 CURRENT APPLICATION NUMBER: US/09/839,894
 CURRENT FILING DATE: 2001-04-20
 PRIOR APPLICATION NUMBER: 60/198,626
 PRIOR FILING DATE: 2000-04-20
 NUMBER OF SEQ. ID NOS: 40
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 32
 LENGTH: 353
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: ETEC Protein Homology Sequence
 US-09-839-894-32

Query Match	13.1%;	Score 247.5;	DB 9;	Length 353;
Best Local Similarity	28.9%;	Pred. No. 6.8e-15;		
Matches	89;	Conservative 42;	Mismatches 126;	Indels 51;
				Gaps 13;

QY	77	NTLNG-ACPTSEINSSSS-----VSGFTNTTLOFTEKRSIKKELDIKGYKOLFK	126
Db	72	NVLGWCWCRSNRNEGGEETHLWVWYAFGAYSIRLRFREDISHAETTL-----	1200
QY	127	SVNCPGSLTNSAHFNC-----NKRAAS-----CASLXLYTPAGELKNPEFGGWTDAITLKL-R	178
Db	121	-----ILLGSVDACGVTNNMAAQQMQRSLKLRIPSELKLIPTSGWATITLVLDY	173
QY	179	VKRRYSFTGYTYTNTITKLTLD--KGNQIWLPOF-KSDARVLDLNLPRPGGTYIGNSV	235
Db	174	LQMGCDPLGISTDTIDLTNLNVDEHAENNAIIFPOFGATPRPVDLNLHRMNASQMSGANL	233

[illegible]

Query Match	6.3%	Score 119.5;	DB 10;	Length 770;
Best Local Similarity	20.4%	Pred. No. 0.0083;		
Matches 83;	Conservative 48;	Mismatches 140;	Indels 135;	Gaps 20

```

0Y 9 TLFFSSVETFAVAVADKIPGDESTITFEGRDNRESSPKHILUNH-----54
Db 249 TIAITSTYIGSFAADLKNNSSSSSSN-----SSNNSTQIUNGKLNKHWITLLIYLINLSAK 304
0Y 55 --TAYSESHTLYDRMTFLCSHNTLNGACPTSENNSSSSVSGETNITLOFTEKRSLLK 112
Db 305 VCLAGYHEKRLHLD-----LLDPYNTLERPVLNESDPLQSLFGLTLMQIIDVDENKQLLV 359
0Y 113 RELQIKGYKQLLEKFSVNCPSGLTINSAHFNCNKMAAGASILYITPAGELKNLPFGCIWD 172
Db 360 TNNVMLK-----LEWMDNKLRMNSTDYGVK-----DLR-IPPHRIWK 395
0Y 173 ATLMKRVKRRYSEI-----YGYTITINITIKLTKDNQIOIWLPO--FKSDARVDNLRLPTGG 226
Db 396 PDVLH-----TNSADEGDEGYTQNNVVR--NNGSC--LYVPGLGIFKSTCKIDITWPF-- 444
0Y 227 GTYIGRNSVDMCF-----YDGYSTNSSSLHTRFQDNNEKS-----DCKEFLRKINDTKE 276
Db 445 ---FDDQCEMKFCSWYIDGF-----QLDIQLQDDEFTGGDISSYVLNENWELGJPGKRNE 466
0Y 277 IAYT-----LSLLAGKSLTPNGTSLNADA 303
Db 497 IYVNCDEPYIDITFAIIRRRRTLYEFFNLIPCVLIASMALGFTLPDPGSEKLSIGVT 556
0Y 304 ASLE-TNNNRITAYTMEPLS--VAV-----LCW-----PGR 331
Db 557 ILSLTYVLNNVAVETMPATSDAVPLMIRYVFLCWLPMILNMRSGR 602

```

RESULT 8
US-10-092-880-10
; Sequence 10, Application US/10092880
; Patent No. US20020164354A1
; GENERAL INFORMATION:

```

: APPLICANT: Barenkamp, Stephen J.
: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
: TITLE OF INVENTION: HAEMOPHILUS
: FILE REFERENCE:
: CURRENT APPLICATION NUMBER: US/10/092,880
: CURRENT FILING DATE: 2002-03-08
: PRIOR APPLICATION NUMBER: 09/155,614
: PRIOR FILING DATE: 1998-09-30
: PRIOR APPLICATION NUMBER: 08/617,697
: PRIOR FILING DATE: 1996-04-01
: PRIOR APPLICATION NUMBER: PCT/US97/04707
: PRIOR FILING DATE: 1997-04-01
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 10
: LENGTH: 1600
: TYPE: PRT
: ORGANISM: Haemophilus influenzae
: US-10-092-880-10

```

Query Match	6.28;	Score 116.5;	DB 9;	Length 1600;
Best Local Similarity	21.68;	Pred. No. 0.044;		
Matches 74; Conservative	60;	Mismatches 130;	Indels 79;	Gaps 19

```

QY 48 HNLINNIITYSESHETLYDMRTFLCISSHHTLGNACPTSENSS- ----SVSGGTINTLT 102
Dd 816 IYNEYSKH -AINSSHNL-----TIIIGNVTLTG-----ENSSSITGININTNKANVTL 862
QY 103 QF-----TEKRSILKRELOIKGKTKOLLFKSVNCPSSLTL-----NSAHFNCN--- 144
Dd 863 QADTSNSNTGLKRRTLTLGNISVEGNLSITGAAMNIVGNLSIAEDSTFKGEISDMLNTGT 922
QY 145 ---KNAASGASLYLYIPAGELKMLPEFGJWDATLKLRRVRRSEYGT- YTIATINIKLTD 200
Dd 923 TETNNGTANINIKGVYKVLGDINK- GG-----LNTITNASGTOKTLINGINT- NE 970
QY 201 KGNIOIWLPOFKSDARVDL- NLRPTGGGTYIGRNSVYDMCFDYGSTNSSLEIFQ--- 255
Dd 971 KGDJINT- --KIKKADAEIOJIGNISOREGNLTJSSDKVNT-----TNDYTIKAGVEGSR 1021
QY 256 -DNPNKSDGFEYLRKIIDDPKEIATVYLSLLIAG- --KSLPTNGTSLNADASALETWMN 311
Dd 1022 SDSEAEANMLTIQ-----TKELKLAGDNLISEFNKAEITANKGSDTLIGNASGGMADAK 1076
QY 312 RITAVTMEISeVYLCMPG- RLOLDAKVENPEAGOTMGINVT 353
Dd 1077 K---VTEDKVDSKISTDGHNVNLTSEVTSINSGSSNAGDNST 1116

```

```

RESULT 9
US-09-912--020-302
Sequence 302, Application US/09912020
Patent No. US20020045592A1
GENERAL INFORMATION:
APPLICANT: Zyskind, Judith
APPLICANT: Ohlsen, Karl L.
APPLICANT: Trawick, John
APPLICANT: Forsyth, R. Allyn
APPLICANT: Froelich, Jamie M.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
TITLE OF INVENTION: ESCHERICHIA COLI
FILE REFERENCE: ELITRA.00IDV1
CURRENT APPLICATION NUMBER: US/09/912,020
PRIORITY FILING DATE: 2001-07-23
PRIORITY APPLICATION NUMBER: 09/492,709
PRIORITY FILING DATE: 2000-01-27
PRIORITY APPLICATION NUMBER: 60/117,405
PRIORITY FILING DATE: 1999-01-27
NUMBER OF SEQ ID NOS: 485
SOFTWARE: FastSeq for Windows Version 3.0

```


SEQ ID NO 302
LENGTH: 2383
TYPE: PRF
ORGANISM: E. COLI
US-09-912-020-302

Query Match 5.8%; Score 109; DB 10; Length 2383;
Best Local Similarity 20.0%; Pred. No. 0.37;
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

QY 13 SSVLFPAVSADKIPDEDTITIFGPRDRNESSPKHILNNHITAY---SESHLTIDRMT 69
DB 762 SAKIATLASNNGVLANEANAATVSVNADEGS---NPINDHTVTPAVLISGATSFNNO 818
QY 70 FLCLSSHNTLNGACPTSENPSSSVSGETINTITLOFTEKRSLLI-----K 112
DB 819 ----TAKTIVNGLA-TFDLKSSSQEDNTVEYTIENGVKOTLIYSPGDSSTAVIDOKSK 873
QY 113 RELQIRKGYKOL-LFKSVNCPGSLTILNSAHFNCKNAASGLYLIYPAGELKNLPPGGIM 171
DB 874 NEVVADGNDSYMVTATVRDAKGNLNDVAVTFVNVSAEAKLSQTEVNSHD-----GIA 926
QY 172 DATLKLARKVRRSEYTGTTITITIKLTKGNQIWLPOPKSDARVDLNRPGGTYIG 231
DB 927 TATLT-----SLKNGDYRYTASVSSGSQANQVNFIDQSTAAULTLV-PSGDITV-- 976
QY 232 RNSVDMCFYDGYSTNS---SLEIRFODNPKSDGKFLRKINDTKELIAYTILLSLAG 287
DB 977 -----TNTAPQYMTATLQDKNGNPLAKKEITFSPND-----VASKFSISNG 1019
QY 288 KSLTPTNGTSL-----NIADAASLETMMNRTTAVTMEISVPLC 327
DB 1020 KGMTDSNGVAIASLFTGLAGTHIMARLANSVSDAQPMTFVADKRAVAVLQTSKAEII 1079
QY 328 WPG--RLQDAKVENPEAGQMGNIIVTFT 355
DB 1080 GNGVDETTTLTATVKDP-SNHVAGITVNF 1108

RESULT 10
US-10-092-880-9
Sequence 9, Application US/10092880
Patent No. US20020164354A1
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
TITLE OF INVENTION: HAEMOPHILUS
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/092,880
CURRENT FILING DATE: 2002-03-08
PRIOR APPLICATION NUMBER: 09/155,614
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 08/617,697
PRIOR FILING DATE: 1996-04-01
PRIOR APPLICATION NUMBER: PCT/US97/04707
PRIOR FILING DATE: 1997-04-01
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 9
LENGTH: 1599
TYPE: PRF
ORGANISM: Haemophilus influenzae
US-10-092-880-9

Query Match 5.4%; Score 101; DB 9; Length 1599;
Best Local Similarity 21.5%; Pred. No. 1.1;
Matches 90; Conservative 52; Mismatches 172; Indels 104; Gaps 18;

QY 21 VSADKIPGDESTINIGPRDRNESSPKHNI-LNNHI-TAYSESHLTIDRMTFLCLSS---- 74
DB* 1077 VTFDKV-----KDSKISTDGHNVITLNSEYKTINGSSNAGNDNSTGLTISAKIV 1124
QY 75 -----SHNTLN-----GACPTSENPSSSVSGETINTITLQ-----FTEKR 108

DB 1125 TVNNVTSKHTINISAAAGNVTTKEGTTINATGVSVEYTAQNTIKNTISQNVYTA 1184
QY 109 SLIKRELQI-----KGYKOLLFKSVNCPGSLTILNSAHFNCKNAASGLI----- 153
DB 1185 NVATTENAVINATSGTVNSTGTGIDKIGIEESTSGNVNI---TASGNTLKVENITQDVT 1241
QY 154 -----YLTIYPAGELKNLPPGGIMDATLKLARKVRRSEYTGTTITITIKLTKGNQIWL 207
DB 1242 VYADAGALTTTGSTISATGNAANTTTKTGIDNGKVESSSGSVTLTATATLAVGISGN 1301
QY 208 LPOFKSDARVDLNRPGGTYIGRNSVDMCFYDGYSTNSSSLEIRFODN---PKSDG 264
DB 1302 TVTTFADSG---KLTFSTVGSTINGTNSVTT-----SSQSGDIEGTISGNTVNTASTGD 1352
QY 265 FYL-RKINDTEIAYTLS-----LLAGKSLTIPNGTSLINADAASLETWNN--RITA 315
DB 1353 LTIGNSAKVEARKNGAATTLTAESGKLTTQYTGSSITSSNGTTLTAKDSSIAGNINANAYTL 1412
QY 316 VTMPEISV-----PVLCPGRQLQ---DAKVENPEAG-----QYMGNIIVTFTPS 358
DB 1413 NTTGTLTTTGDSKINATSGTLTINAKDAKLDGASGDRIVNATNASSGNTAKTSS 1470

RESULT 11
US-09-839-894-6
Sequence 6, Application US/09839894
Patent No. US20020176868A1
GENERAL INFORMATION:
APPLICANT: Alboum, Zeev
APPLICANT: Barry, Eileen M.
APPLICANT: Levine, Myron M.
TITLE OF INVENTION: Isolation and characterization of the
TITLE OF INVENTION: CSA OPERON
FILE REFERENCE: USFMD.006A
CURRENT APPLICATION NUMBER: US/09/839,894
CURRENT FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: 60/198,626
PRIOR FILING DATE: 2000-04-20
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 867
TYPE: PRF
ORGANISM: E. COLI
US-09-839-894-6

Query Match 5.0%; Score 95; DB 9; Length 867;
Best Local Similarity 23.6%; Pred. No. 1.6;
Matches 47; Conservative 24; Mismatches 56; Indels 72; Gaps 10;

QY 125 FKSVCPSGLTILNSAH-----FNCKNAASGASLYIYPAGELKNLPPGCTDATLKLRY 179
DB 435 YOSISTYDGFSLSFYHNDKRVNDCGRNYNAGWS-GCYESYASLSIPILG-WTSTL----- 488
QY 180 KRYSEF-----YGFYINITIKLTKGNQIWL-LPOFKSDARVDLNRP--- 223
DB 489 --GYSDTISESYKHHIILSEYGFNONIT-----KGTORWQLTSTSLKMDYFMPEAI 541
QY 224 -----TGGGYI-----GNSYDMCFYDGYSTNSSS-- 249
DB 542 GIYNSRQRLQTDKGGIYSTLTRASHENS LNNGYVNSRKYSSNELFVDDYMTSTNG 601
QY 250 -----LEIRFODNPKSDGK 264
DB 602 DYHEVAMREKNNRHNNAEGR 620

RESULT 12
US-09-797-862-33
Sequence 33, Application US/09797862
Patent No. US20020102276A1

```
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRN
; ORGANISM: Haemophilus influenzae
US-09-797-862-33

Query Match          5.0%; Score 94.5; DB 10; Length 2353;
Best Local Similarity 19.7%; Pred. No. 7.6;
Matches 61; Conservative 47; Mismatches 117; Indels 85; Gaps 14;

QY 78 TLNACPTSENSSSVSGETNITLQTEKRSILKRELQIKGYQLLEKSYNC---PSG 133
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 806 TIGNTPT---GGTATPKVNT--STADGLNFAKETADASGSKNVYLKGIATTLTBSA 860
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 134 LTLNSAHNCNKA---ASGASLYLTPA-----GELKNLPFGIMDATLKRKRKSE 185
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 861 -GAKSSHVDLVNDATKKSNAASIEDVLRAGWNIGNGNNVDYATYDV-----NFTDD 913
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 186 TYGTYTITITIKLTDK-----NIOIMLPQFSDARVDLNL---RPTG 225
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 914 STGTYTYVTKAGKAGKADVIGAKTSYIKOHNKRLFQKDLKANNAGATYSEDDGKDTG 973
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 226 GGTYYIGNSVDMCTYDGT-----STNSSLETRFDNN----- 258
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 974 TGLVTAKTVIDAVNKGWRTVEGATETGATAVNAGNAEVTSGTSYVFNKGNATYATV 1033
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 259 PKSDGKFLR-----KINDTKETAYTSL-LLAGKSLTPNGSLN---IADAA 304
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 1034 SKDGNINVKIDVNVGDLKIGDDKKIYADTTTLTVTGKAVPAGANSVNNKKLVNAE 1093
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 305 SLETNMNRIT 314
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 1094 GLATALNNLS 1103
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

UNT 13
09-924-154-14
; Sequence 14, Application US/09924154
; Patent No. US20020127241A1
; GENERAL INFORMATION:
; APPLICANT: Natrum, David L.
; APPLICANT: Sim, Kim L.
; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
; FILE REFERENCE: 05213-0465 43170-262105
; CURRENT APPLICATION NUMBER: US/09/924,154
; CURRENT FILING DATE: 2001-08-07
; PRIOR APPLICATION NUMBER: US 60/223,525
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1143
; TYPE: PRN
; ORGANISM: Mammalian
US-09-924-154-14

Query Match          4.9%; Score 91.5; DB 10; Length 1143;
Best Local Similarity 20.0%; Pred. No. 5;
Matches 84; Conservative 49; Mismatches 152; Indels 135; Gaps 21;
```

```
QY 22 SADKIPDESIYN-IFGPRDRNESSPKHNLIN-----HTAYSESHLYDRMT--FL 71
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 100 SYDIIPPSYVRNDKFNLSLENEDNSGNTSNPNANTSELSICKDNQYFIQKRTHLFA 159
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 72 CLSSHTNLNACPTSENSSSVSGETNITLQTEKRSILKRELQIKGYQLLEKSYNCP 131
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 160 CGIKRKSIIKWC--RENSEKTIWC--VPDRKIOLCIANFLNSRLETMEKFEIELISVNT 216
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 132 SGLTLNSAHNCNKAASGASLYLTPAGELKN-----LPRGIMDATLKL 178
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 217 AKLY-----NKNCKDPSIF---CNELRNSFSDFRNSFTGDDMDFGNTD-RVKGY 264
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 179 VKRRYSEYGYTYI---NIT-----IKLTDRGN-----IQI 206
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 265 INKFSDIYKKNVEKNELNNIKKEWMEKNKANIMNMHMYVNHKNISKCALIIPAEPPQINL 324
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 207 WLPOFKSD---ARVDLNRPTGGGTYIGRNSVDKCF-----YDGKSTNSSS--- 249
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 325 WIKEMNENFLMEKKRLFLNLIK---DKCVENKKYEAFCGGRCPSSSYTSFMKRSKTOME 380
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 250 -----LEIRFDNNPKS--DGKEYLRKINDTKETAYTSLILLAGKSL 290
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 381 VLNLVKKKNSGVDKNNFLNDLFEKKNKNDLDFEKKKEYDLDCCRYATYTI--KSF 437
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 291 TPTNGTSLNIADAASL-----ETNNRITAYTMPETISYVLCWPGRLQ 333
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 438 -LNGPAKNVDIASQINVMDLRGFCGNCYKSNKNSNCTGFTN--KPPGCEPRRQ 492
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 14
US-09-944-160-12
; Sequence 12, Application US/09944160
; Patent No. US20020174452A1
; GENERAL INFORMATION:
; APPLICANT: Lewis, No. US20020174452A1man
; APPLICANT: Davin, Laurence
; APPLICANT: .. Huang, Ning
; TITLE OF INVENTION: Monocot Seeds with Increased Lignan
; TITLE OF INVENTION: Content
; FILE REFERENCE: WSUR117983
; CURRENT APPLICATION NUMBER: US/09/944,160
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 60/230,632
; PRIOR FILING DATE: 2000-09-07
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 564
; TYPE: PRN
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: laccase amino acid sequence from plasmid paf1245
US-09-944-160-12

Query Match          4.8%; Score 90.5; DB 9; Length 564;
Best Local Similarity 22.0%; Pred. No. 2.3;
Matches 80; Conservative 57; Mismatches 132; Indels 95; Gaps 22;

QY 1 MNKILFI-----FTLFFSVLFTFAVSADKI---PDESTINTFGGRDNNESSPKHNL 51
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 230 MNTTFESIASHSVTVVGSDAAYTKPKLSDYITISPG-QYIDFL---QANQPSHYM 264
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 52 NNHTAYSESHLYDRMTFLCLSSHTNLNACPTSENSSSVSG--ETNITLQTEK-R 108
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 285 AARAYAVAGN---EDNTTTAIIIRYKG-NYTAAPS--PSPFNLPGFMDTNASVAFYRLR 338
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 109 SLIKRELQIKGY---QLLEK-SVN--CESSGLTLNSAHNCNKAASGASLYLTPAG 160
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 339 SLGNKNKYPVDVKNVTKLILFTFSINLTPCPN---NSCAGPFNERPRASVNNITTEVP-- 392
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 161 ELKMLPFGIMDATLKLVRKRKRYSEYGT---YTNITIKITDKGNQIMLPQFKSDARY 217
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

Db 333 -----PTAILOAYQ-RIRNVYSNNPSPNPFENYTSIIIPR---DLMRPQNGTEVKV 443
 Oy 218 DLNLAPRTGGGTYIGNSVDKCFYDGYSTNSSLEIRPDNNPKSGKRYLKRINDTDEI 277
 Db 444 -----LKNSTVEIVFOGTN-----ILGIDH----- 465
 Oy 278 AVTLSLIAGKSLPTNGTSLINADASLETNMNRITAVTPEISVPVLCMPGRLOLDAK 337
 Db 466 -----PIHLHGOSFYV-CWGLGNFNNAIDPLNLYNDPPLNNTIAPVSGMTA---VRFK 517
 Oy 338 VEMP 341
 Db 518 ASNP 521

RESULT 15

US-10-011-588-29
 ; Sequence 29, Application US/10011588
 ; Patent No. US2002016872A1
 GENERAL INFORMATION:
 APPLICANT: Smith, Leonard
 ; APPLICANT: Jensen, Melody
 ; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
 ; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
 ; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
 ; FILE REFERENCE: A34796 067252.013
 ; CURRENT APPLICATION NUMBER: US/10/011,588
 ; CURRENT FILING DATE: 2002-03-29
 ; PRIOR APPLICATION NUMBER: 09/910,186
 ; PRIOR FILING DATE: 2001-07-20
 ; PRIOR APPLICATION NUMBER: 09/611,419
 ; PRIOR FILING DATE: 2000-07-06
 ; PRIOR APPLICATION NUMBER: 60/246,744
 ; PRIOR FILING DATE: 2000-11-06
 ; PRIOR APPLICATION NUMBER: 60/311,966
 ; PRIOR FILING DATE: 2001-08-09
 ; NUMBER OF SEQ ID NOS: 47
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 29
 ; LENGTH: 811
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:28
 US-10-011-588-29

Query Match 4.7%; Score 89.5; DB 9; Length 811;

Best Local Similarity 18.4%; Pred. No. 4.7;
 Matches 76; Conservative 60; Mismatches 112; Indels 165; Gaps 19;

Oy 2 NKILFIFLFFSSVLETPFAVSADKIPGDESTITFGPRDRNESSPKHNLNNHITAYSES 61
 Db 351 NELQIFLFEFYAKIYN--VONRKI---YLSNYTP-----VTANILD 388
 Oy 62 HTLYDRMFLCL--SSHTLNGACPTSENPSSSVSGFTNITLOFTEKRSILKRELQIKG 119
 Db 389 DNVDIÖNGFNIPKSNLVLTMGONLSRNPARKVNP-NMLYLTFCFCSLNTKTLDC-- 445
 Oy 120 YKQLFKSVNCPGSLTNSAHFNCNKNAASGSLYIIPAGELKNLPF-GGIWDATLKLK 178
 Db 446 -RELVKNT-----DLPIGDISDYKTDIF 469
 Oy 179 VKRRYSEYGT--YTINITI-----KLTDKNIQIWLPOFKSDARVDLNLPTGGGTY 229
 Db 470 LRKDNIEETEVLYYPDNVSVDVILSKNTSEHQDLPLPSIDSESEI-----L 518
 Oy 230 IGRNSVDKCFYDGYSTNSSLEIRPDNNPKSGDKFYL--RKINDTKEIAYTLSTL---- 283
 Db 519 PGENOV---FIDNRKQNDYL-----NSYYLESQKLSDNVEDFTTRISIERAL 564
 Oy 284 -----LLAGK-----SLTPTN 294
 Db 565 DNSAKVYTYTFPLANKVNAVÖGGFLMWANDVVEDFTTNILRKDTLIDKISDVSAIIPYI 624

Oy 295 GSNLIADASLETNMNRITAVT-----MPETSVPLCMRGRLOLDAKVE 339
 Db 625 GPALNINSVR-RGNFTEAFAVTGVTILLEAFPEFTIPAL--GAFVITYSKVQ 673

Search completed: December 6, 2002, 18:19:11
 Job time : 18 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2002, 17:55:05 ; Search time 17 Seconds
(without alignments)
624.804 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886

Sequence: 1 MNKILFIPTLFSSVLEFFFA.....EAGGYMGNINVTTPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1536.5	81.5	360	2	US-08-483-101-17
2	988.5	52.4	363	2	US-08-483-101-16
3	927.5	49.2	364	2	US-08-483-101-5
4	116.5	6.2	1600	2	US-08-617-697-10
5	114	6.0	650	3	US-08-362-525-2
6	111.5	5.9	1005	4	US-09-206-942-41
7	111.5	5.9	1011	4	US-09-206-942-39
8	111.5	5.9	1529	2	US-08-728-470-10
9	111.5	5.9	1529	2	US-08-719-641-10
10	107	5.7	1095	4	US-09-206-942-45
11	107	5.7	1101	4	US-09-206-942-43
12	106	5.6	671	2	US-08-737-716-13
13	101	5.4	915	4	US-09-206-942-35
14	101	5.4	1222	4	US-09-206-942-37
15	101	5.4	1228	4	US-09-206-942-34
16	101	5.4	1338	2	US-08-728-470-9
17	101	5.4	1338	4	US-08-719-641-9
18	101	5.4	1599	2	US-08-617-697-9
19	100.5	5.3	1004	4	US-09-206-942-57
20	100.5	5.3	1010	4	US-09-206-942-55
21	99.5	5.3	969	4	US-09-206-942-32
22	99.5	5.3	975	4	US-09-206-942-30
23	99	5.2	992	4	US-09-206-942-61
24	99	5.2	998	4	US-09-206-942-59
25	99	5.2	1848	4	US-08-296-791-6
26	99	5.2	1848	5	PCT-US95-10661A-6
27	97	5.1	977	4	US-09-206-942-53

28	97	5.1	983	4	US-09-206-942-51	Sequence 51, Appl
29	94.5	5.0	1912	1	US-08-409-995-4	Sequence 4, Appl
30	94.5	5.0	1912	3	US-08-685-467-4	Sequence 4, Appl
31	94.5	5.0	2353	4	US-09-377-155-33	Sequence 33, Appl
32	94.5	5.0	2353	4	US-08-913-942-4	Sequence 33, Appl
33	94.5	5.0	2353	4	US-09-669-974-33	Sequence 33, Appl
34	94.5	5.0	2354	4	US-09-268-347-47	Sequence 36, Appl
35	94.5	5.0	2411	4	US-09-268-347-36	Sequence 36, Appl
36	93.5	5.0	901	4	US-09-134-001C-5351	Sequence 5351, Ap
37	93	4.9	1167	1	US-08-485-568A-6	Sequence 6, Appl
38	93	4.9	1167	2	US-08-590-554A-6	Sequence 6, Appl
39	93	4.9	1167	2	US-09-184-223-6	Sequence 6, Appl
40	92.5	4.9	1073	4	US-09-206-942-49	Sequence 49, Appl
41	92.5	4.9	1079	4	US-09-206-942-47	Sequence 47, Appl
42	91.5	4.9	869	2	US-08-483-101-15	Sequence 15, Appl
43	91	4.8	1168	1	US-08-620-717A-9	Sequence 9, Appl
44	90.5	4.8	1328	3	US-08-781-891-76	Sequence 76, Appl
45	90	4.8	569	4	US-09-173-053-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-483-101-17
; Sequence 17, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80503
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Feiber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; US-08-483-101-17
Query Match 81.5%; Score 1536.5; DB 2; Length 360;
Best Local Similarity 80.1%; Pred. No. 8e-150;
Matches 289; Conservative 33; Mismatches 38; Indels 1; Gaps 1;
QY 1 MNKILFIPTLFSSVLEFFFAVSADKIPGDESITNIFGPRDNSSPKHILNNHTVAYSE 60
DB 1 MNKILFIPTLFSSVLEFFFAVSADKIPGDESITNIFGPRDNSSPKHILNNHTVAYSE 60

QY	61	SHTYDNRHTEFLCLESNHTLNACQTESNPSSVSGENITLQPTERRSLIKRELQJGK	120
Dp	61	SHHXLDNRHSFLCLESQNTLNACPSDPAFGATIDGETNITLQPTERRSLIKRELQJGK	120
QY	121	KÖLLEKFSVNCPSGTLTNSAHFNCKNKAASGASLYLYTPAGELNKLPGGIWDATLKLVRK	180
Dp	121	KÖPFEKFNANCPSKALNLSHPOCRREOASGATLSLYIPAGELNKLPGGVWNVKLKNVK	180
QY	181	RRYSEFGTYVINTIKTLTDKGNIOIQLPQKRSAPARYDLNLRPTGGGTYIGRNSVDMCFY	240
Dp	181	RRYOTTYGTYINTITVNLTDKGNIOIQLPQKRSARADNLNLRPTGGGTYIGRNSVDMCFY	240
QY	241	DGYTNSSSLEIRQDNNPKSDGKFYLRKINDTKETAYVLSLLLAGKSLTPPTNGTSLNT	300
Dp	241	DGYTNSSSLEIRQDNNPKSDGKFYLRKINDTSKELEYVLTLSLLLAGKSLTPPTNGALNT	300
QY	301	ADAASLEFNMRITAYVTPMEISYVVLCPMGRLODAKVEKNPEAGQYMGNIYVTFPSSQ	360
Dp	301	NTFASLEFNMRITAYVTPMEISYVVLCPMGRLODAKVKNPPEAGEYSGLINVTFTPSSS	359
	361	L 361	
	360	L 360	

RESULT 2
 US-08-483-101-16
 : Sequence 16, Application US/08483101
 : Patent No. 5932715
 : GENERAL INFORMATION:
 : APPLICANT: Scott, June R.
 : APPLICANT: Froehlich, Barbara
 : APPLICANT: Caron, Judy
 : TITLE OF INVENTION: CS2 Proteins and Coding Sequences
 : NUMBER OF SEQUENCES: 17
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Greenlee and Winner, P.C.
 : STREET: 5370 Manhattan Circle, Suite 201
 : CITY: Boulder
 : STATE: Colorado
 : COUNTRY: US
 : ZIP: 80303
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/483,101
 : FILING DATE: 07-JUN-1995
 : CLASSIFICATION: 424
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Ferber, Donna M.
 : REGISTRATION NUMBER: 33878
 : REFERENCE/DOCKET NUMBER: 6-95
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (303) 499-8080
 : TELEFAX: (303) 499-8089
 : INFORMATION FOR SEQ. ID NO: 16:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 363 amino acids
 : TYPE: amino acid
 : STRANDEDNESS:
 : TOPOLOGY: unknown
 : MOLECULE TYPE: protein
 : HYPOTHETICAL: NO
 : US-08-483-101-16

Query Match	52.48	Score 988.5	DB 2	Length 363
Best Local Similarity	55.08	Pred. No. 2.2e-93		
Matches 204	Conservative 53	Mismatches 93	Indels 21	Gaps 11

[illegible]

```

1      RESULT 3
2      ; Sequence 5, Application US/08483101
3      ; Patent No. 5932715
4      ; GENERAL INFORMATION:
5      ; APPLICANT: Scott, June R.
6      ; APPLICANT: Froehlich, Barbara
7      ; APPLICANT: Caron, Judy
8      ; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
9      ; NUMBER OF SEQUENCES: 17
10     ; CORRESPONDENCE ADDRESS:
11     ; ADDRESSEE: Greenlee and Winner, P.C.
12     ; STREET: 5370 Manhattan Circle, Suite 201
13     ; CITY: Boulder
14     ; STATE: Colorado
15     ; COUNTRY: US
16     ; ZIP: 80303
17     ; COMPUTER READABLE FORM:
18     ; MEDIUM TYPE: Floppy disk
19     ; COMPUTER: IBM PC compatible
20     ; OPERATING SYSTEM: PC-DOS/MS-DOS
21     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: US/08/483,101
24     ; FILING DATE: 07-JUN-1995
25     ; CLASSIFICATION: 424
26     ; ATTORNEY/AGENT INFORMATION:
27     ; NAME: Ferber, Donna M.
28     ; REGISTRATION NUMBER: 33878
29     ; REFERENCE/DOCKET NUMBER: 6-95
30     ; TELECOMMUNICATION INFORMATION:
31     ; TELEPHONE: (303) 499-8080
32     ; TELEFAX: (303) 499-8089
33     ; INFORMATION FOR SEQ ID NO: 5:
34     ; SEQUENCE CHARACTERISTICS:
35     ; LENGTH: 364 amino acids
36     ; TYPE: amino acid
37     ; TOPOLOGY: linear
38     ; MOLECULE TYPE: protein
39     ;
40     ; US-08-483-101-5

```

Query Match	49.28;	Score 927.5;	DB 2;	Length 364;
Best Local Similarity	48.18;	Pred. No. 4.3e-87;		
Matches 1/6; Conservative	65;	Mismatches 118;	Indels 7;	Gaps 3

QY 1 MNKILFIPLFFSSVLTFAVSADKIPGDESTINFGPRDRN--ESSPKHNLNNHTAY 58
1 LKKVIFVLSMFLCSQVYQGSMTNVEAGSINKTESIGPDRBAASYPNHYIFHEHVAGY 60
QY 59 SESHLYDRMFLCLSSHNTLNGACPTSENPSSSVSGETNTTLOFTEKRSILKRELQIK 118
61 NNDHSLFDMFLCMSDASGKACPTGENSKS--QGFTNIKILFTEKKSILARTLNK 118
QY 119 GKOLLEFSVNC---PSGLTLNSAFNCKNAASGASLYLYIPAGELKLPFGGIMDATL 175
119 GKRLFLYSDRCIHYVDKNNLSHTVKGCSFTREVDFTLYIPOGELIDLLGGIWEAL 178
QY 176 KLRVRRSEYETGYTNTIKTLTKGNIOIWLPOFKSDARVDLNLRPFGSTYIGRNSV 235
179 ELRVRRHNDYNGTYKNTVTLDTKGNIQVTPKFSHPRIDLNLREPKNGKYSQSNVL 238
QY 236 DMCIFYDGYSTNSSLERFQDNNPKSDGKFYLRLKINDTKELIAYTLSSLAKSLPTNG 295
239 EMLCYDGYSTHQSISLIERFQDSDQGTGNMEYNLIKTEGELPKLIPYKLSLLGGREFYPNG 298
QY 236 TSLNLTADALETNNRITAYMPLEISVPLVLCWPGRLQLDKAVENDEAGQYMGNIWFT 355
239 EAFITNDPSLEPINNRKISVSLPQISIPVLCWPNALTFMSELNPEAGEYSGIINVTFT 358
QY 356 PSSOTL 361
DB 359 PSSSL 364

RESULT 4
US-08-617-697-10
Sequence 10, Application us/08617697
Patent No. 5977336

GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-10

Query Match 6.2%; Score 116.5; DB 2; Length 1600;
Best Local Similarity 21.6%; Pred. No. 0.016;
Matches 74; Conservative 60; Mismatches 130; Indels 79; Gaps 19;

QY 48 HNLNNHTAYSESHLYDRMFLCLSSHNTLNGACPTSENPSS---SVSGETNTTL 102
DB 816 YNEYSKH-AINSSHL-----TILGNAVTLGG-----ENSSSITGININTNNANYTL 862
QY 103 QF-----TEKRSILKRELQIKGYKOLFFRSVNCPSGLT-----NSAHFNCN--- 144
DB 863 QADTSNNTGKLRITLTLGNISVEGNLSITGANAINVGLSLAEDSTFKEGASDLNLTIG 922
QY 145 ---KNAASGASLYLYIPAGELNLNLPFGGIMDATLKLVRKRYSEYGT-YTINITIKLTD 200
DB 923 FTNNGTANINIKGVYKLGIDNNK--GG-----LNTTNASGTQKTIINGNIT---NE 970
QY 201 KGNIOIWLPOFKSDARVDL--NLRPFGSTYIGRNSVDMCFYDGYSTNSSLERFQ--- 255
DB 971 KEDLNI--KNIKADAEIOIGNISOKEGNLTITSSDKVNI-----TWOITKAGVEGGR 1021
QY 256 -DNNPKSDGKFYLRLKINDTKELIAYTLSSLAG---KSLPTNGSLNLTADALETNN 311
DB 1022 SDSEENANLTIQ-----TKELKLAGDLNLSGFNKAELITAKNGSDLTIGNAGGNADAK 1076
QY 312 RITAVTPEISVPLVLCWP-RLQLDKAVENDEAGQYMGNIWFT 353
DB 1077 K---VTFDKVKDSKISTDGHNVTLNSEVYKTSNGSSNAGDNST 1116

RESULT 5
US-08-362-525-2
Sequence 2, Application US/08362525
Patent No. 6027910

GENERAL INFORMATION:
APPLICANT: KLIS, FRANCISCUS M.
APPLICANT: SCHREUDER, MARTEN P.
APPLICANT: TOSCHKA, HOLSER Y.
APPLICANT: VERRIPS, CORNELIS T.
TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D. C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,525
FILING DATE: 04-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 92202080.5
FILING DATE: 08-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 92203899.7
FILING DATE: 14-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/01763
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.

Oy 276 EIAVTLSELLAG---KSLPTNGTSLNIADASLETNNNRITAVTMPEISVPLCMGRL 332
Db 569 ELLTNINISGPNKAEITADNSDLIGKASDSNSNAKOITFDVKXSKIS--AGNNIV 646
Oy 333 OLDKVENPEA-----GQYMGNTVTFPPSSQTL 361
Db 647 TLNSKVEITSNSDGTGNSDDNNIGLITTSARKDVTV 681

RESULT 8
US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/728,470

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Beirstresser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-633

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1529 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-728-470-10

Query Match 5.9%; Score 111.5; DB 2; Length 1529;
Best Local Similarity 20.7%; Pred. No. 0.048;
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

Oy 48 HNIILNHTAASESTLYLDKRTFLCLSSHNLTNGACPTSENPSSS-----SVSGETNITL 102

Db 744 YNEYSKH--AINSSHNL-----TTLGNAVTLGG-----ENSSSSITGINITNKNANVTL 790

Oy 103 QF-----TEKSLIKRELQIKGYKQLLFKSVNCPGLTILNSAHFNCNKNASGASLY 154

Db 791 QADPSNMTGLKRTLTLGNISVEGNLSLTGANNANIVGNLSI--AEDSTFKGAS----- 843

Oy 155 LYIPAGELKNLPFGCIW--DATLKLKRVKRRYSFYGYTV-----INITIKLT----- 199

Db 844 -----DNLNITGTFNNNGTANINIKQYVKLGIDINNGGLNITTNASGOTKTIING 895
Oy 200 ----DKGNIQITLPOFKDARVDL--NRPPTGGGYTIGRNSVDKCFYGYCYSNSSLER 253
Db 896 NITNKGDLNT--KNIKADAETIGGNISQKEGNLTJISSDKVNI-----TNQITIKAG 946
Oy 254 FO----DNNPKSDGKFYRKINDPKEIAYTLSELLAG---KSLPTNGTSLNIADASL 306
Db 947 VEGGRSDSSEANANLTIQ-----TKELKLAGDLNISGFNKAETIRAKKGSDLTITNAGSG 1001
Oy 307 ETNNRITAVTMPEISVPLCMG--RIQLDAKVENPEAGQYMGNTV 353
Db 1002 NADAKK---VTEDKVKDSKISTDGHNTVTLNSEVKTSGNSNAGNDNST 1046

RESULT 9
US-08-719-641-10
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/719,641

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9205704.1

FILING DATE: 16-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Beirstresser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-625

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1529 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-719-641-10

Query Match 5.9%; Score 111.5; DB 4; Length 1529;
Best Local Similarity 20.7%; Pred. No. 0.048;
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

Oy 48 HNIILNHTAASESTLYLDKRTFLCLSSHNLTNGACPTSENPSSS-----SVSGETNITL 102

Db 744 YNEYSKH--AINSSHNL-----TTLGNAVTLGG-----ENSSSSITGINITNKNANVTL 790


```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCF/NL95/00170
: FILING DATE: 12-MAY-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: EP 94201353.3
: FILING DATE: 12-MAY-1994
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 671 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus faecalis
: IMMEDIATE SOURCE:
: CLONE: Fig.5a (S. faecalis)
: US-08-737-716-13

Query Match          5.6%; Score 106; DB 2; Length 671;
Best Local Similarity 23.6%; Pred. No. 0.046;
Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;

QY 41 RNESSPKNLT-LNNHITAYSESHLYDRMTFLCLSSHNTLNGACPTSENPSSSV----- 94
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 309 RYATDPSTYNAKLNNVITAY--NLTDYTPSSSGGNTGGTVNPGTGSNNQSGTNTYTYVK 366
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 95 SGET--NTTLPF---TEKRSI--IKRELQIKGYKQLFKSVNCPSSGTLNSAHFNCKN 146
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 367 SGGTLNKIAAOYGVSVANLRSWNGISGDLIFVGQKLIYKKA--SGNTGGSGNGSGNN 423
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 147 AASGASLYIYPAGELKMLPGGIWDATLKLVRKRYSEYGTYYTINITIKLTDGNIQI 206
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 424 -QSGTNTYTYVKSGDTLN-----KIAAQG-----VTV-----ANLRS 455
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 WLPOKRSARVDNL-----RPTGGTYIGRNSVDMCFDYSTNSSSLEIRFQDNN 258
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 456 W-----NGISGDLIFVGQKLIYKKGTSNT-----GSSNGS-----NNN 491
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 259 PKSDGKFLRKINDTKELA--YTLS-----LLAGSLRPTNGTSLINDA 303
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 492 OSGTNTYTYINSGDTLNIAOYGVSVANLRSWNGISGDLIFAGOKLIYKKGTSNTGGS 551
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 ASLETNNMR 312
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 552 SNGSGNNNO 560
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
: -09-206-942-35
: Sequence 35, Application US/09206942
: Patent No. 6432669
: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M.
: APPLICANT: Yang, Yan-ping
: APPLICANT: Klein, Michel H.
: TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
: TITLE OF INVENTION: Molecular Weight Proteins
: FILE REFERENCE: 1038-861 MIS:jdb
: CURRENT APPLICATION NUMBER: US/09/206, 942
: CURRENT FILING DATE: 1998-12-08
: EARLIER APPLICATION NUMBER: 09/167, 568
: EARLIER FILING DATE: 1998-10-07
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 35
: LENGTH: 915
: TYPE: PRT
: ORGANISM: Haemophilus Influenzae
: US-09-206-942-35

Query Match          5.4%; Score 101; DB 4; Length 915;
Best Local Similarity 21.1%; Pred. No. 0.25;
Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;
```

```

QY 21 VSADKIPGDESTINIFGR-----DRNESSPKH-----NILN-----NHITAYS 59
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 252 ISSDKVNITERITIKAGVNGDSNSNEATSAULTTKTELTNDLNSGENKAEITAKD 311
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 ESH-FLYD-----RMFP-----LCSSHN--TLNGACPTSENPSSSSVSGEINI 100
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 312 NSNLTIGNSDAGNTDAKKVTFSSNVKDSKISASDHNVTLNSKVEISGDTDEGDNNT 371
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 101 TLQFTKRSLLIKRELQIKGYKOL--LFKSYNCPBSGLTLNSAHFNCKNNAASGASLYLYI 157
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 372 GLTITAKWVTYVNN--NITSHKVTNITASENVTTKAGTTIN-----ATTGSVEYTA 419
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 158 PAGELKNLPFGGIMDATLKLVRKRYSEYGTYYTINITIKLTDKGNIOIWLPOKRSARV 217
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 420 KTGDIK-----GGI-----EBSNGVNI----- 437
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 218 DLNLRTGGGTYYI-----GRNSVDMCFYDGYSTNSSSLEIRFQDNNPKSDGKFLRKIN 271
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 438 -----TASGDTLNSNITGQN-VTVAASGAVTTTKGSTINATGNANITTK--TGEIN 488
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 272 DDTKEIAVTLSELLAGKSLPTNGTSLNIADAASLETNMNRITAVTPEISVPVLCWPCR 331
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 489 GEVKSASGNVNITASGNTLNSNITGQWVTYVAN-----SGAITTEGST----- 533
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 332 LQIDAKVENPEAGQYMGNNIN-----VFTPSSQTL 361
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 534 --INATGDANITTOTGTINGKVESSSGSVTLIATGOTL 570
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
: US-09-206-942-37
: Sequence 37, Application US/09206942
: Patent No. 6432669
: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M.
: APPLICANT: Yang, Yan-ping
: APPLICANT: Klein, Michel H.
: TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
: TITLE OF INVENTION: Molecular Weight Proteins
: FILE REFERENCE: 1038-861 MIS:jdb
: CURRENT APPLICATION NUMBER: US/09/206, 942
: CURRENT FILING DATE: 1998-12-08
: EARLIER APPLICATION NUMBER: 09/167, 568
: EARLIER FILING DATE: 1998-10-07
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 37
: LENGTH: 1222
: TYPE: PRT
: ORGANISM: Haemophilus Influenzae
: US-09-206-942-37

Query Match          5.4%; Score 101; DB 4; Length 1222;
Best Local Similarity 21.1%; Pred. No. 0.4;
Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;
```

```

Oy 218 DLNLRPTGGGYYI-----GNSYDMCFYDQSYNSSLSEIRFODNNPKDKGFEYLRKN 271
Db 745 -----TASGPTLVNSNTGQN-VTAAASGAVTTTGTSINATGGANAITTK-TGEIN 795
Oy 272 DDTKEIATLTLGAGSLTPTNCTSLINDAASLEINMNITVTPMPELSVPLCLMPGR 332
Db 796 GEVSASGNNVITTSAGNTLVNSNTGQNVITYAN-----SGAITTBGST----- 840
Oy 332 LQDDAKYENPEAGQTMGNIN-----YTFPSSQTL 361
Db 841 --INATGGDANITTTGTGNGINKVESSGGSVYLLATGGTL 877

```

```

RESULT 15
US-09-206-942-34
; Sequence 34, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS.1b
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 34
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-34

```

Query Match	5.48;	Score 101;	DB 4;	Length 1228;
Best Local Similarity	21.18;	Pred NO. 0.4;		
Matches 84;	Conservative 48;	Mismatches 129;	Indels 138;	Gaps 19;

[illegible]

Search completed: December 6, 2002, 18:18:30
Job time : 22 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2002, 17:20:37 ; Search time 38 Seconds

(without alignments)
1265.881 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886

Sequence: 1 MNKILFFITLFFSSVLTFFFA.....EAGGYMGNNVITFFPSQTL 361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : A.Geneseq_101002.*

1: /SID2/gcgdata/geneseq/genesep-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/genesep-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/genesep-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/genesep-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/genesep-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/genesep-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/genesep-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/genesep-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/genesep-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/genesep-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/genesep-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/genesep-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/genesep-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/genesep-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/genesep-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/genesep-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/genesep-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/genesep-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/genesep-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/genesep-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/genesep-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/genesep-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/genesep-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1886	100.0	361	23	AAM50343
2	927.5	49.2	364	20	AAV22326
3	275.5	14.6	359	22	AAB45919
4	119.5	6.3	770	21	AAV50814
5	114	6.0	650	15	AAR47575
6	111.5	5.9	1005	21	AAB01833
7	111.5	5.9	1011	21	AAB01832
8	110.5	5.9	1529	14	AAR41732
9	110.5	5.9	1601	18	AAM30292
10	109	5.8	2383	21	AAB15945

11	107	5.7	1095	21	AAB01835
12	107	5.7	1101	21	AAB01834
13	106	5.6	671	17	AAR85290
14	106	5.6	1222	21	AAB01830
15	106	5.6	1228	21	AAB01828
16	105	5.6	598	23	AAB47313
17	104.5	5.5	1440	23	AAB54801
18	101	5.4	1338	14	AAR41731
19	101	5.4	1598	18	AAM30291
20	100.5	5.3	1004	21	AAB01841
21	100.5	5.3	1010	21	AAB01840
22	99.5	5.3	969	21	AAB01827
23	99.5	5.3	975	21	AAB01826
24	99	5.2	839	23	AAB98211
25	99	5.2	992	21	AAB01843
26	99	5.2	998	21	AAB01842
27	97	5.1	418	21	AAG30907
28	97	5.1	454	21	AAG30906
29	97	5.1	957	21	AAB01839
30	97	5.1	963	21	AAB01838
31	96	5.1	430	21	AAG37667
32	96	5.1	513	21	AAG39277
33	96	5.1	631	21	AAG39276
34	96	5.1	683	21	AAG39275
35	95.5	5.1	825	21	AAB25547
36	95	5.0	867	23	AAM50341
37	94.5	5.0	567	20	AAV37170
38	94.5	5.0	665	22	AAB68244
39	94.5	5.0	2353	17	AAR99393
40	94.5	5.0	2411	21	AAB23860
41	94	5.0	394	21	AAV37668
42	94	5.0	443	23	AAU91513
43	93.5	5.0	715	20	AAV05625
44	93.5	5.0	901	23	ABP40506
45	93.5	5.0	1464	22	ABP71111

ALIGNMENTS

RESULT 1	
AAM50343	
AAAM50343 standard; Protein; 361 AA.	
XX	
AC	AAAM50343;
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	EPEC CS4 pilus CsaE tip associated protein.
XX	
KW	CS4 pilus; enterotoxigenic; EPEC; csa operon; CsaA; fimbrial;
KW	vaccine; diarrhoea; antibacterial; antidiarrhetic.
XX	
OS	Escherichia coli.
XX	
FT	Key
FT	Peptide
FT	Protein
FT	Location/Qualifiers
FT	1..23
FT	/label= Signal_peptide
FT	24..361
FT	/label= Mature_protein
PN	W0200181582-A2.
PD	01-NOV-2001.
XX	
PF	20-APR-2001; 2001WO-US12914.
XX	
PR	20-APR-2000; 2000US-198686P.
XX	
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Altbaum Z, Levine MM, Barry EM;
XX	

Haemophilus influe
Haemophilus influe
Streptococcus faec
H. influenzae scra
Haemophilus influe
Listeria monocytog
Lactococcus lactis
High molecular wei
Non-typeable Haemo
Haemophilus influe
Haemophilus influe
Haemophilus influe
Haemophilus influe
Chlamydia polypept
Haemophilus influe
Haemophilus influe
Arabidopsis thalia
Arabidopsis thalia
Haemophilus influe
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Eucalyptus grandis
EPEC CS4 pilus Csa
Amino acid sequenc
Drosophila melanog
Haemophilus adhesi
Haemophilus influe
OspC-Pko/OspA-B31/
HIV-1 group O 1sol
Staphylococcus epi
Drosophila melanog

DR WPI; 2002-049280/06.
XX N-PSDB; AAI70763, AAI70780.
PT New nucleotide sequence, useful as immunogenic agent for generating
PT immune response against recombinant product of the operon, comprises
PT csa operon which encodes enterotoxigenic Escherichia coli-CS4 pili -
XX
XX
PS Claim 10; Page 59; 81pp; English.
XX
XX The present sequence is that of the tip associated protein CsaE
CC of enterotoxigenic Escherichia coli (ETEC) strain E11881A. CsaE is
CC encoded by the csa gene (see AAI70763) of the E. coli E11881A csa
CC operon. This operon includes 5 contiguous genes, csaA-csaE, which
CC encode the synthetase of ETEC-CS4 pili. It has been expressed in
CC attenuated Shigella strain CVD1204 guinea, constructing the Shigella
CC expressing CS4 fimbriae vaccine strain CVD1204 (PGA2-CS4). The
CC CsaE protein has a calculated mol.wt. of 40102.4 and a theoretical
CC pI of 8.74. It shows homology to similar proteins from other ETEC
CC fimbriae. Recombinant CsaA-CsaE polypeptides are used in claimed
CC immunogenic compositions to generate an immune response in a
CC subject. These prevent ETEC colonisation, and hence protect
CC against diarrhoea.
XX
XX Sequence 361 AA:
SQ
Query Match 100.0%; Score 1886; DB 23; Length 361;
Best Local Similarity 100.0%; Pred. No. 4e-169;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNKILFIFLTFESSVLTFEFAVSADKIPGDESITNIFGPRDRN--ESSPKHNLNNHTAY 60
DB 1 MNKILFIFLTFESSVLTFEFAVSADKIPGDESITNIFGPRDRN--ESSPKHNLNNHTAY 60
QY 61 SHTLYDRMTFLCLSSHNTLNGACPTSENPSSSVSGETNITLQFTKRSLLKRELQIKGY 120
DB 61 SHTLYDRMTFLCLSSHNTLNGACPTSENPSSSVSGETNITLQFTKRSLLKRELQIKGY 120
QY 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLTPAGELKMLPFGGIMDAFLKLRVK 180
DB 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLTPAGELKMLPFGGIMDAFLKLRVK 180
QY 181 RRYSETGTNTINTIKLTDKGNIOIWLPOKSPARVDLNLPTGGGTYIGRNSVDMEFY 240
DB 181 RRYSETGTNTINTIKLTDKGNIOIWLPOKSPARVDLNLPTGGGTYIGRNSVDMEFY 240
QY 241 DGYSTNSSLEIRFODNNPKSDGKFLRKINDTKELAYTLSSLAKSLPTNGSLNI 300
DB 241 DGYSTNSSLEIRFODNNPKSDGKFLRKINDTKELAYTLSSLAKSLPTNGSLNI 300
QY 301 ADAASLETNNRITAVTMPETISVPVLCWPGRLQDAKAVENPAGQYMGNTINVTTPSSQT 360
DB 301 ADAASLETNNRITAVTMPETISVPVLCWPGRLQDAKAVENPAGQYMGNTINVTTPSSQT 360
QY 361 L 361
DB 361 L 361
RESULT 2
ID AAY22326 standard; Protein: 364 AA.
XX AAY22326;
XX
XX 22-SEP-1999 (first entry)
XX
XX Piliin protein Csd.
XX
XX CS2 gene cluster; Cota; Cota; Cota; Cota; Piliin protein; immunogen;
XX enterotoxigenic E. coli; human upper intestine; diarrhoeal disease;
XX
XX Escherichia coli.

XX
XX US5932715-A.
XX
XX 03-AUG-1999.
XX
XX 07-JUN-1995; 9505-0483101.
XX
XX 07-JUN-1995; 9505-0483101.
XX
XX (UYEM-) UNIV EMORY.
XX
XX Caron J, Froehlich B, Scott JR;
XX
XX WPI; 1999-443623/37.
XX
XX N-PSDB; AAX84848.
XX
XX Isolated nucleic acids encoding Escherichia coli CS2 pili proteins
XX useful for vaccinating against diarrhoeal diseases caused by
XX Escherichia coli
XX
XX Claim 3; Column 45-48; 35pp; English.
XX
XX This sequence represents a CS2 pili protein of the invention, encoded by
XX (which also encodes Cota, Cota, and Cota). CS2 pili are long
XX proteinaceous molecules thought to mediate attachment of enterotoxigenic
XX E. coli (ETEC) to and/or promote colonisation of the human upper
XX intestine. The CS2 gene cluster may be used to produce immunogens for
XX vaccinating patients against diarrhoeal diseases caused by ETEC bacteria.
XX This type of enteric infection is a major cause of death among infants in
XX developing countries and in immunocompromised (e.g. Acquired Immune
XX Deficiency Syndrome (AIDS)) or elderly adults. The vaccine comprises more
XX than 1 antigenic determinant (epitopes) from more than 1 pili type to be
XX effective against more than 1 type of ETEC infection.
XX
XX Sequence 364 AA:
SQ
Query Match 49.2%; Score 927.5; DB 20; Length 364;
Best Local Similarity 48.1%; Pred. No. 9.1e-79;
Matches 176; Conservative 65; Mismatches 118; Indels 7; Gaps 3;
QY 1 MNKILFIFLTFESSVLTFEFAVSADKIPGDESITNIFGPRDRN--ESSPKHNLNNHTAY 58
DB 1 LKVIIFVLSMFLCSQVYQGSWHTNVEAGSINKTISGPIRDSAAASPAHIFHEHAGY 60
QY 59 SESHTLYDRMTFLCLSSHNTLNGACPTSENPSSSVSGETNITLQFTKRSLLKRELQIK 118
DB 61 NKDHSLEDRMTFLCLSSHNTLNGACPTSENPSSSVSGETNITLQFTKRSLLKRELQIK 118
QY 119 GYKQLEKSVNCPGSLTNSAHFNCKNAASGASLYLTPAGELKMLPFGGIMDAFLK 175
DB 119 GYKQLEKSVNCPGSLTNSAHFNCKNAASGASLYLTPAGELKMLPFGGIMDAFLK 175
QY 176 KLRVRRYSETGTNTINTIKLTDKGNIOIWLPOKSPARVDLNLPTGGGTYIGRNSV 235
DB 176 KLRVRRYSETGTNTINTIKLTDKGNIOIWLPOKSPARVDLNLPTGGGTYIGRNSV 235
QY 236 DMCIFYDGYSTNSSLEIRFODNNPKSDGKFLRKINDTKELAYTLSSLAKSLPTNG 295
DB 236 DMCIFYDGYSTNSSLEIRFODNNPKSDGKFLRKINDTKELAYTLSSLAKSLPTNG 295
QY 296 TSLNIAADAASLETNNRITAVTMPETISVPVLCWPGRLQDAKAVENPAGQYMGNTIN 355
DB 296 TSLNIAADAASLETNNRITAVTMPETISVPVLCWPGRLQDAKAVENPAGQYMGNTIN 355
QY 356 PSSQL 361
DB 356 PSSQL 361
RESULT 3
ID AAB45919 standard; Protein: 359 AA.
XX AAB45919
XX

AC AAB45919;
XX
DT 23-MAR-2001 (first entry)
XX
DE S. enterica serovar Typhi tcfD fimbrial subunit protein.
XX
KW Fimbrial protein; saf; tcf; vaccine; gene therapy; immunization;
KW tcf insert; detection.
XX
OS Salmoneella typhi.
XX
PN WO200073336-A1.
PD
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-SE01079.
XX
PR 28-MAY-1999; 99SE-0001961.
XX
PA (ACTI-) ACTIVE BIOTECH AB.
XX
FI Folkesson A, Normark S, Loeffdahl S;
XX
DR WPI; 2001-061512/07.
DR N-PSDB; AAC82926.
XX
PT Fimbriae proteins of Salmoneella enterica subspecies I bacteria, useful
PT for producing vaccines against the bacterial subspecies and for
XX detecting the bacteria -
XX
PS Disclosure; Page 72-73; 77pp; English.

	Query Match	14.6%	Score 275.5	DB 22	Length 359	
	Best Local Similarity	29.7%	Pred. No.2.6e-17			
	Match 90:	Conservative	48;	Mismatches 130;	Indels 35;	Gaps 11.
OY	70	FICLSHNTLNACPTSENPSSSVSGENITLQFEKRSILKRELQIGYKQLLEFSVN	129			
		::::: : : : : : : : : :				
Db	79	WCRSNRNEBACETHLVMMYAAGAVSKIRPREQISHAEITL-----ILLGSVR	131			
OY	130	--CPGSLTNSAHFNCNKNAAS--GASLYELIPAGELLNLPFGGIWDATLKL-RVKRRY	183			
		: :				
Db	132	DACYTCV-----INMAAACQMSRLSKLIPSELAKIPTSGWKRTVLVDIQLQMG	183			
OY	184	SETGYTYINITIKLTD--KGNIQIMLPQF-KSDARVDNLNRPITGGGTIGRNSVDMCFY	240			
		: : : : : :				
Db	184	DDPLGTSTIDILNTLDPHEANAALTFPQGFTATPRVDDLNRMANMQSGRANDMCLY	243			
OY	241	DGYSTNSSSELRFODNNPKSDGKFYLKRIINDTYEIAVTLSSLLAGKSFLPNGTSLNI	300			
		::: : : : : : : : :				
Db	244	DG-GVKARSLQAKIREGSNKSCTGFQIKSDSADT-IDVAVMNYNGGRSIIPVTRGVEEST	300			
OY	301	AQAASLETWMNIIIVATMEISVPVLCWPGRQLQDAK---VENPEAQGMGINIVTFPS	357			
		::: : : : : : : : :				
Db	301	DVVDKAAFR----PVYLEGQRQAVRCVVPPLLTLLTPQPIIEKRGEGLFYTMIMG	355			

QY	358	SQT	360
		:11	
Db	356	TQT	358

RESULT 4
AAV50814
ID AAV50814 standard; Protein; 770 AA

DT 17-FEB-2000 (first entry)

D. melanogaster acetylcholine receptor protein from clone Da7.

KW Acetylcholine receptor; nicotinic; insect; insecticide; screening;
neurotransmission; plant protection agent; conductance; AChR.

OS *Drosophila melanogaster*

PN DE19819829-A1

PD 11-NOV-1999.

PF 04-MAY-1998; 98DE-1019829.

PR 04-MAY-1998; 98DE-1019829.

PA (FARB) BAYER AG.

PI Adamczewski M, Oellers N, Schulte T;
.....

DR WPI; 2000-014207/02.

XX
XX

PT insects, used to identify potential

PS Example 1a; Page 12-14; 26pp; German.
xy

CC This invention describes a novel nucleic acid (NA) encoding a nicotinic
CC acetyl-choline receptor (I) from insects which can be used as an
CC insecticide. The NA can be used to transform a host cell (H) with a transgene (T) to produce a transgenic host cell (T) which can be used to produce a transgenic insect (TI) which can be used to control insect pests.

CC (also vectors containing it, its regulatory regions, and antibodies directed against it) needed to be used to generate a library of

CC protection agents that alter conductance of AChR, potentially useful as

CC formation of functionally related AChR in insects. (I) are also used to

CC recombinant production of (II). This sequence represents an

SQ Sequence 770 AA;

Query Match	6.3%	Score 119.5	DB 21	Length 770
Best Local Similarity	20.4%	Pred. No. 0.041		
Matches 83	Conservative 48	Mismatches 140	Indels 135	Gaps 20

```

QY 9 TLFSSVLTFAVSADKIPGDESITNIFGPRDRNESSPKHNLNNH----- 54
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dh 340 MATFSGVCGEMOLKNGSGGGGGG---CGNNGCTOTNCTNINUKSCTRTTVIINCAV 304

```

00 55 - - T T A V S E S U M T V D B M T E I C I S U N N T I N C A C P B T S E N D S C S E V C S E M N T I O E T E K P C I T V 1 1 2

305 VOT ACYHEKPII HD- - - - - I DRYNTI EBYOVI NEECHT CI SECT TI MOTIDVEKNOITV 350

00 113 PETOTKCYKOTI EKSVNDSCTI NSAHENCNKNA SCACI VITBACEI KNP DECCIED 177

Db 360 TNNVTK-----TENNDMNTPWNTSDYGCYK-----DLP-TPBHRITWK 395

0v 173 A7IKI BVKBPVSET - - - - VGTVTNTITKI TDKGNTOIWI BO - - EKSDARVDI NI BPTGC 326

11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533

Db 396 PVLV-----YNSADEGDTGYQTVNVR--NNGSC-LYVPEIGKSTCKIDITWFP---- 444

QY 227 GYVIGRNSVDKCF---YDGYSTNSSLEIRFQDNPKS-----DGKFLRKINDTKE 276

Db 445 ---FDDQREKMFSGSWTYGCF-----QDLDQLQDERGSGISSVYLVNGEMLGCPKKRNE 496

QY 277 IAYT-----LALLAGKSLPTNGTSLNIADA 303

Db 497 IYVNCCEPEYIDITFAIIRRLTYFFENLIIPCVLIASMALGFTLPPDSCGKLSLGYT 556

QY 304 ASLE-TNNMRITAVTPEIS--VPV-----LCW-----PGR 331

Db 557 ILLSLTVFLNMVAETMPATSDAIPMIRIVFLCWLPMILRMSRPR 602

RESULT 5
AAR47575
ID AAR47575 standard; Protein; 650 AA.

XX AAR47575;
XX
XX 19-JUL-1994 (first entry)
XX
XX Alpha-agglutinin of *Saccharomyces cerevisiae*.
XX
XX Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;
XX KM Major cell wall protein; glycosyl-phosphatidyl-inositol;
XX KM anchoring protein; alpha factor; alpha-agglutinin; invertase;
XX KM inulinase; alpha-amylase; *Saccharomyces cerevisiae*; catalysis.
XX KM enzymatic process; fermentation; biodegradation; catalysis.
XX OS *Saccharomyces cerevisiae*.
XX
XX PN WO9401567-A.
XX PD 20-JAN-1994.
XX
XX PF 07-JUL-1993; 93WO-EP01763.
XX
XX PR 08-JUL-1992; 92EP-0202080.
XX PR 14-DEC-1992; 92EP-0203899.
XX
XX PA (UNITL) UNILEVER NV.
XX PA (UNITL) UNILEVER PLC.
XX PI K11s FM, Schreuder MP, Toschka H, Verrips CT;
XX WPI: 1994-035071/04.
XX N-PSDB: AAO54012.
XX
XX PT Immobilisation of enzymes to microbial cell wall - by prodn. of
XX PT fusion protein of enzyme linked to anchoring protein
XX
XX PS Example 1; Page 32-39; 99pp; English.
XX
XX CC The alpha-agglutinin is used in a method to immobilise enzymes to a
XX CC microbial cell wall. The coding sequence is used in the production
XX CC of a recombinant polynucleotide which comprises a structural gene
XX CC encoding a protein with catalytic activity and at least part of a
XX CC gene encoding at least the C-terminus of a protein capable of
XX CC anchoring in a eukaryotic or prokaryotic cell wall. The anchoring
XX CC fragment or protein is selected from alpha agglutinin, AGA 1, FLO 1,
XX CC major cell wall protein of lower eukaryotes or a proteinase of
XX CC lactic acid bacteria. The recombinant polynucleotide preferably
XX CC also comprises a sequence encoding a signal peptide to ensure
XX CC secretion of the expressed product. The signal peptide is
XX CC preferably derived from glycosyl-phosphatidyl-inositol, anchoring
XX CC protein, alpha factor, alpha-agglutinin, invertase or inulinase,
XX CC alpha-amylase of *Bacillus* or proteinases of lactic acid bacteria.
XX CC The host microorganism can be used for performing enzymatic
XX CC processes on an industrial scale.
XX
XX SQ Sequence 650 AA:
XX

Query Match 6.0%; Score 114; DB 15; Length 650;
Best Local Similarity 20.9%; Pred. No. 0.1;
Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

QY 4 ILFFITLFFSSVL-----TFPA-----VSADKTPG-----DESTINIFCPDRNE--- 43

Db 8 ILMFSLALASAININDIFSSNLEITPLANKOPDDGWTATPFDFTADASIREGDEFTL 67

QY 44 SSP---KHNLINNHITA-----YSESHTYDRMTFLC-----LSSHNTLN 80

Db 68 SMPHYRIRIKLNSSQTATITSLADGTEAFKCYVSOQAHYILENTTFTCTAONDLSNTITD 127

QY 81 GACPTEENPSSSSVSGETNITLOFTEKRSILKRELQIKGYKOLLFSVNC-PSGLTLNSA 139

Db 128 GSITFSLNFSDDGSSYEYEL-----ENAKFEKSGPMLVKGNDQVNFMDPAALTEVNF 182

QY 140 HFNCNKNNAASGASLYL--YIPAGELKNLPGGIWDATLKLRYKRRYSFYGYTTINITTK 197

Db 183 HSGRSTGYGSEFSEYHLGMYCPNGVF---LGG---TERI-----DYDSSNNNVLD- 225

QY 198 LMDKGNIOI-----WLPQEKSDARVDLNRPTGGGYIGRN---SVDMCFYDGC--- 242

Db 226 --DCSSVOYSSNDFNDMMFPOSYNPTNADV-----TCFESNIMTLDEKLYDGEML 275

QY 243 -----YSTNSSLE--IRFQDN--PKSDGKF 265

Db 276 WYNALQSLPANVTIDHALFQYTCIDTIANNTYATQFTTREFIYYQGRNLGTASAKSS 335

QY 266 YLRKINDTKET--AYTLS-----LLAGKSLPTNGTSLNIADA 304

Db 336 FISTTTDLTSLNTSAYSGSISTVETGNRTSEVISHVVTSTKSLPATTSLSLTIAQTS 395

QY 305 SLETMNN-----RITAYTMEISVPVLCWPGRL 332

Db 396 IYSTDSNITVGDIDHTTSEVISDVEFTISRETASTVAAPTSTTGWTGAM 444

RESULT 6
AAB01833
ID AAB01833 standard; Protein; 1005 AA.

XX AAB01833;
XX
XX AC AAB01833;
XX
XX DT 11-SEP-2000 (first entry)
XX
XX DE Haemophilus influenzae strain K21 mature HMW2A protein, SEQ ID NO:41.
XX
XX KM Mature HMW protein; hmw gene; hmwA2; high molecular weight;
XX KM non-typable Haemophilus influenzae; NTH; non-encapsulated;
XX KM recombinant production; Escherichia coli; antibacterial; vaccine;
XX KM human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX KM detection; diagnosis.
XX
XX OS Haemophilus influenzae strain K21.
XX OS
XX PN WO200020609-A2.
XX
XX PD 13-APR-2000.
XX
XX PF 07-OCT-1999; 99WO-CA00938.
XX
XX PR 07-OCT-1998; 98US-0167568.
XX PR 08-DEC-1998; 98US-0206942.
XX
XX PA (CONN-) CONNAUGHT LAB LTD.
XX
XX PI Loosemore SM, Yang Y, Klein MH;
XX
XX DR WPI: 2000-303789/26.
XX DR N-PSDB: AAA52182.
XX
XX PT Nucleic acid molecule for producing recombinant high molecular weight

PT proteins of Haemophilus which are used as a vaccine to provide
 PT protection against Haemophilus induced diseases in humans -
 XX
 PS Claim 8; Fig 21A-O; 307pp; English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwaBC operon from a non-typable (non-encapsulated) H. influenzae (NTNI). Most HMW-expressing NTNI strains contain two hmw gene clusters termed hm1AaC and hm2AaBC. Each hmwaBC operon comprises hmwa, hmwb and hmwc genes. The hmwa genes encode the structural HMWA proteins and the hmwb and hmwc genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwaBC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwa genes (AA521175-552198) and HMWA proteins (AA801824-B01849) from the non-typable H. influenzae strains Joyce, KI, K21, ICPC2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the non-typable strains of Haemophilus via hybridisation reactions. The present sequence represents a mature HMWA protein from a non-typable strain of H. influenzae.

SQ Sequence 1005 AA;

Query Match	5.9%;	Score 111.5;	DB 21;	Length 1005;
Best Local Similarity	20.3%;	Pred. No. 0.34;		

[illegible]

RESULT 7
AAB01832
ID AAB01832 standard; Protein; 1011 AA.

AC	AAB01832;
XX	
DT	11-SEP-2000 (first entry)
XX	
DE	Haemophilus influenzae strain K21 HmWZA protein, SEQ ID NO:39
XX	

KW HmV protein; hmv gene; hmvA1; hmvA2; high molecular weight;
KW non-tyeable Haemophilus influenzae; NHI; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis
KW detection; diagnosis.

OS Haemophilus influenzae strain K21

PN WO200020609-A2

PD 13-APR-2000.

PF 07-OCT-1999; 99WO-CA00938.

PR 07-OCT-1998; 98US-0167568.

[illegible]

24 XX

XX
XX
T
C
C
C
C
C
C
C
C
C

DR N-PSDB; AAA52181.

PT Nucleic acid mole

PT protection against Haemophilus induced diseases in humans -

PS Claim 12; Fig 21A-O; 307pp; English

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in *Escherichia coli*. The expression construct used to effect recombinant expression comprises a promoter functional in *E. coli* (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typable (non-encapsulated) *H. influenzae* (NTHi). Most HMW-expressing NTHi strains contain two hmw gene clusters termed hmwIABC and hmwVABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwa genes (AA55217-5-A52198) and HMWA proteins (AA001824-B01849) from the non-typable *H. influenzae* strains Joyce, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant *H. influenzae* HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by *H. influenzae* (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typable strains of Haemophilus via hybridisation reactions. The present sequence represents an HMWA protein from a non-typable strain of *H. influenzae*.

SQ Sequence 1011 AA;

Query Match	5.9%	Score 111.5;	DB 21;	Length 1011;
Best Local Similarity	20.3%;	Pred. No. 0.35;		
Matches 68; Conservative	65;	Mismatches 147;	Indels 55;	Gaps 14

```

OY      60  ESHLTYDMLTCLSSNHTLWAGACPTSPNBPSSSSVSGETINTLOFTEKRS-LIKRELOIK 118
Db      369  EKNALEFSTHNLTLIGANTVLTIGENSSSSIKGINININSTANTLOAHAGTSHLDKE---- 4244
OY      119  GYKOLLFESVNCPSGLT-LSAHFNCNKNAASGASLYLYIPAGEL-KNLPFGIW--DA 173
Db      425  --RTLLTGVNVSGLNLTIGSNAHIDGMLSTAESAKE-----QCKTNMNNINTCTFPNNG 477
OY      174  TLKLRVKRYSSETGYTINTIKLTD-----KGNL-----QIMLPQFSDARVD 218

```

```

Db 478 TADINIKQGVKLOGDITNNGNINITTNASVNOKTIINGNITNKGDNLINIKANAIEQ 537
QY 219 L--NLRPTEGGTYIGRNSVDMCFYDGYSTNSSLEIRFQDNPKSD--GKFYLRKINDDK 275
Db 538 IGNINISQKEGNLTISDKIYI-----TKRIEKADTDQGNDSGVSANANITIKTK 588
QY 276 EIAVYLSLLAG--KSLPTNGTSLNIADAASLETNMNRITAVTMEISVPVLCWPGRL 332
Db 589 ELTLTDLNLTNSGFKAEITAKDNSDLIGKASSDNSNAKQITFDKVKSKIS--AGNHNV 646
QY 333 QLDKAYENPEA-----GQYMGNIINVTFTPSQTL 361
Db 647 TLNSKVETNSDGSSTGNGSDDNIGILTISAKDYTV 681

```

RESULT 8

AA041732
ID AAR41732 standard; Protein; 1529 AA.

AA041732;
AC

26-APR-1994 (first entry)
XX

High molecular weight protein 4 (HMW4).
DE

HMW; high molecular weight protein; virus; vaccine; influenza;
epitope; immunity; haemophilus influenzae.
KW
XX

Haemophilus influenzae.
OS

WO9319090-A.
-PN

30-SEP-1993.
PD

16-MAR-1993; 93WO-US02166.
PF

16-MAR-1992; 92GB-0005704.
PR

(BARE/) BARENKAMP S J.
(INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PA

Barenkamp SJ;
PI

WPI; 1993-320683/40.
XX

N-PSDB; AAQ49511.
DR

High molecular weight surface proteins - of non-typeable
haemophilus which exhibit immunogenic properties
PT

Claim 6; Figure 10; 100pp; English.
XX

The isolation and purification of the high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines.
CC

Sequence 1529 AA;
XX

Query Match 5.9%; Score 110.5; DB 14; Length 1529;
Best Local Similarity 20.7%; Pred. No. 0.8;
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

```

QY 48 HNLINHTAYSESHLYDMFTCLSSHTLNGACPTSPSS-----SVSGETNITL 102
Db 744 YNEYSKH--AINSSHNL-----TLAGNVTLGG-----ENSSSITGINITNKAVTL 790
QY 103 QF-----TEKRSLIKRELOIKYKQLLFKSVNCPGSLTNSAHFNCKNNAAGASLY 154
Db 791 QADTSNNGKRTLTIGNISVEGNISLGTANANITGNLSI--AEDSTFKGEAS----- 843
QY 155 LVIPEGLKLNLPGGIW--DATLKRKRYSETYGT--INTIKLT----- 199

```

```

Db 844 -----DNLNITGTEFTNNGTANINIKQGVKLOGDIINKGGLNITTNASGTOKTLING 895
QY 200 ----DKGNIDIMLPQFSDARVLD--NLRPTEGGTYIGRNSVDMCFYDGYSTNSSLEIR 253
Db 896 NITNEKGDLENI--KNIKADEIQLGNGISQKEGNLTISDKVNI-----TMOITIKAG 946
QY 254 FO-----DNNPKSDGKFLRKINDTKEIAVYLSLLAG--KSLPTNGTSLNIADAASL 306
Db 947 VEGGRSDSSAEANANLTIQ-----TKELKLAGDLNLTNSGFKAEITAKNGSDLTIGNASGG 1001
QY 307 ETMNRITAVTMEISVPVLCWPG--RLQLDKAYENPEAGQYMGNIINVT 353
Db 1002 MADAKK---VTFDKVKDKSISTDGHVNTLNSVETKSNSSNAGNDNST 1046

```

RESULT 9

AA030292
ID AAW30292 standard; Protein; 1601 AA.

AAW30292;
AC

14-APR-1998 (first entry)
XX

Non-typeable Haemophilus high mol.wt. surface protein HMW4.
DE

Non-typeable Haemophilus; high molecular weight surface protein;
KW HMW4; Immunogen; vaccine; otitis media.
XX

Haemophilus influenzae strain 5.
OS

Key Location/Qualifiers
FH Misc-difference 372

FT /note="encoded by TCT"
FT Misc-difference 400

FT /note="encoded by AAT"
FT

WO9736914-A1.
PN

09-OCT-1997.
XX

01-APR-1997; 97WO-US04707.
PF

01-APR-1996; 96US-0617697.
PR

(BARE/) BARENKAMP S J.
PA

Barenkamp SJ;
PI

WPI; 1997-503038/46.
XX

N-PSDB; AAT90993.
DR

High molecular weight proteins of non-typeable Haemophilus
influenzae - useful for vaccine production
PT

Claim 1; Page 97-102; 183pp; English.
PS

This protein comprises the high molecular weight surface protein
CC HMW4 (123 kDa) of non-typeable Haemophilus influenzae strain 5 that
CC has the immunological ability to protect against disease caused by
CC a non-typeable Haemophilus strain and is characterised by at least
CC one surface-exposed B-cell epitope that is recognised by monoclonal
CC antibody A06. The HMW4 amino acid sequence was deduced from an
CC isolated hmw4 gene (see AAT90993). HMW1 (see AAW30293), HMW2 (see
CC AAW30294) and HMW3 (see AAW30291) have also been identified. A
CC conjugate comprising HMW4 linked to an antigen, hapten or
CC polysaccharide, and a synthetic peptide of 6-150 amino acids
CC corresponding to at least protective epitope of HMW4 are also
CC claimed. HMW proteins, conjugates and peptides can be used in
CC vaccines, as immunogens for preparation of antibodies and as
CC antigens for detection of these antibodies.

Sequence 1601 AA;
SQ

DR WPI: 2000-303789/26.
XX N-PSDB: AAA52184.
XX Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
XX Claim 8: Fig 22A-P: 307pp: English.
XX
XX The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwaBC operon from a non-typable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmwaABC and hmwa2ABC. Each hmwaBC operon comprises hmwaA,
CC hmwb and hmwc genes. The hmwa genes encode the structural HMWA proteins
CC and the hmwb and hmwc genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwaBC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwa genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae
CC strains Joyce, K1, K21, LCD2C, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMWA protein from a non-typable
CC strain of H. influenzae.
XX
XX
SQ Sequence 1095 AA:
Query Match 5.7%; Score 107; DB 21; Length 1095;
Best Local Similarity 24.6%; Pred. No. 1;
Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;
QY 45 SPKHNILNHNITAYSESHTYD-----RMFLCISNHTLNGACPTSENPSSSS 93
DB 281 SIKPPIYSN---VHDGNHTLFNGNVSVLGGDVNFHFNASSSNHMTGVIKSONFNAS 337
QY 94 VSGETNITLOFTEKRLIKRELQIKGYKQLFKSVNCPGLTNSAHFNCKNKAASGASL 153
DB 338 GS-----SLRF-----KSEGSTRAF---TIESDLTNAAGNLSLQVAGIDG 378
QY 154 YLYIPAGELKNLFPFGIWDATLKLVRKRRSEYGYTYT---NITIKLTKGNITQIWL 209
DB 379 NLQSKLVAANKNITFEG---GNITLAADKPIETIKGNTIVEGANVTLRSANYGNDKRAL 434
QY 210 QEISDARVDNLAPRTGGGTIGRN-----SYDQCFYDGYSTNSSSLEIRQDNPKSD-- 262
DB 435 SINGANTNKNGLVYTGSAINIEKNLTVEGSAKFLANPNSYFNSGSL----FDQGKSNIS 490
QY 263 ---GKFLRKINDTKEIAYTLS-----LLAGSKLTPNGSTNSIAD 302
DB 491 IAKGGAHFKNIN-NITSLNITTTNSDSAYRTIIEG-NITNSNG-DLNIITD 536
RESULT 12
AAB01834
ID AAB01834 standard; Protein; 1101 AA.
XX
XX AAB01834:
AC
AC AAB01834:
AC
AC 11-SEP-2000 (first entry)
XX
XX DE Haemophilus influenzae strain LCD2C HMWA protein, SEQ ID NO:43.

KM HMW protein; hmw gene; hmwa1; hmwa2; high molecular weight;
KM non-typable Haemophilus influenzae; NTHi; non-encapsulated;
KM recombinant production; Escherichia coli; antibacterial; vaccine;
KM human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KM detection; diagnosis.
XX
XX Haemophilus influenzae strain LCD2C.
XX
XX MO200020609-A2.
XX
XX 13-APR-2000.
XX
XX 07-OCT-1999; 99WC-CA00938.
XX
XX 07-OCT-1998; 98US-0167568.
XX 08-DEC-1998; 98US-0206942.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Yang Y, Klein MH;
PI
PI
DR WPI: 2000-303789/26.
XX N-PSDB: AAA52183.
XX
XX Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
XX Claim 12: Fig 22A-P: 307pp: English.
XX
XX The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwaBC operon from a non-typable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmwaABC and hmwa2ABC. Each hmwaBC operon comprises hmwaA,
CC hmwb and hmwc genes. The hmwa genes encode the structural HMWA proteins
CC and the hmwb and hmwc genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwaBC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwa genes (AAA52175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae
CC strains Joyce, K1, K21, LCD2C, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents an HMWA protein from a non-typable strain of
CC H. influenzae.
XX
XX
SQ Sequence 1101 AA:
Query Match 5.7%; Score 107; DB 21; Length 1101;
Best Local Similarity 24.6%; Pred. No. 1;
Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;
QY 45 SPKHNILNHNITAYSESHTYD-----RMFLCISNHTLNGACPTSENPSSSS 93
DB 287 SIKPPIYSN---VHDGNHTLFNGNVSVLGGDVNFHFNASSSNHMTGVIKSONFNAS 343
QY 94 VSGETNITLOFTEKRLIKRELQIKGYKQLFKSVNCPGLTNSAHFNCKNKAASGASL 153
DB 344 GS-----SLRF-----KSEGSTRAF---TIESDLTNAAGNLSLQVAGIDG 384
QY 154 YLYIPAGELKNLFPFGIWDATLKLVRKRRSEYGYTYT---NITIKLTKGNITQIWL 209

```

Db 385 NLQSLVANKNITFEG---GNITLADKKPIEIKGNITVEKAGNVLRSANYGDKSAL- 440
OY 210 QPKSDARVDLNRPTGGGTYIGRN-----SYDMCFYDGYSTNSSSEIRFQDNPKSD-- 262
Db 441 STRGNATKNGNLVTGSAINIEKNLTVEGSAKFLANPNSEFNVSGL----FDQGRKSNLS 496
OY 263 ---GKFEYLRKINDDTKEIAVTLG-----LLAGKSLTFPTNGSLNIAD 302
Db 497 IAKGGAHFKPDIN-NTKSLNITTTMSDSAYRTIIEG-NITNSNG-DLNIITD 542

RESULT 13
AAR85290
ID AAR85290 standard; Protein; 671 AA.
XX
AC AAR85290;
XX
DT 02-APR-1996 (first entry)
XX
XX Streptococcus faecalis autolysin.
XX
KM Lysin; autolysis; culture; lactic acid bacteria; fermentation;
KM cheese; foodstuffs; induction.
XX
OS Streptococcus faecalis.
XX
PN W09531561-A1.
XX
PD 23-NOV-1995.
XX
PF 12-MAY-1995; 95WO-NL00170.
XX
PR 12-MAY-1994; 94EP-0201353.
XX
PA (UNIL ) QUEST INT BV.
XX
PI Buist G, Kok J, Ledebor AM, Venema G;
XX
XX WPI; 1996-010946/01.
XX
PT Lysis of a culture of lactic acid bacteria in, e.g. cheese
PT production - by in situ prodn. of an auto-lysin, regulated by an
PT inducible promoter.
XX
XX
PS Disclosure: Page 66-69; 103pp; English.
XX
XX In situ production of a homologous autolysin or a heterologous
XX autolysin from a food grade Gram positive bacteria, can be used in
XX a process for the lysis of a culture of lactic acid bacteria. The
XX process can be used in the manufacture of products containing
XX cultures of lactic acid bacteria e.g. cheese, where the culture is
XX lysed following the completion of fermentation. The enhanced
XX induction of the autolysin is performed some hours after the
XX fermentation is finished. No extra lysis needs to be added and the
XX lysis does not need to be isolated or encapsulated. The time of
XX lysis can be precisely controlled. This is the Streptococcus
XX faecalis autolysin.
XX
SQ Sequence 671 AA;

Query Match 5.6%; Score 106; DB 17; Length 671;
Best local Similarity 23.6%; Pred. No. 0.62;
Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;
OY 41 RNSSPKHNI-LNNHITAVSEHTLYDRMTFLCLSSHNLTNGACPTSENPSSSV----- 94
Db 309 RYATDPSYNKLNKNVITAY--NLTYDTPSSGGTGGTVNPGTGSNNQSGTNYTYVK 366
OY 95 SGT--NITLQF-----TKRSL--IKRELQIKYKOLLFRSVNCPGGLTLNSHFNCNK 146
Db 367 SGTGLTKIAAQGVSVANLRSMNGISGDLLEVGQKLIYKGA--SGNTGSGGSGGNNN 423
OY 147 AASGASLYIYPAGEIKNL.PFGIMDATLKLVRKRYSETGYTITITIKLTDKNQIQI 206

```

```

Db 424 -QSGTNTYTYVSGDFLN-----KIAQYG-----VTV-----ANLNS 455
OY 207 WLPQFKSDARVDLNL-----RPTGGGTYIGRNSVDMCFYDGYSTNSSSEIRFQDN 258
Db 456 W-----NGISGDLIFVGOKLIYKGTSGNT-----GGSNGGS-----NNN 491
OY 259 PKSDGKFEYLRKINDDTKEIA--YTLG-----LLAGKSLTFPTNGSLNIAD 303
Db 492 QSGTNTYTYIKSGDYLTKIAQGVSVANLRSMNGISGDLIFRQKLIYKGTSGNTGGS 551
OY 304 ASLETNMMNR 312
Db 552 SNGGSNNQ 560

RESULT 14
AAB01830
ID AAB01830 standard; Protein; 1222 AA.
XX
AC AAB01830;
XX
DT 11-SEP-2000 (first entry)
XX
XX H. influenzae strain K1 mature full-length HMW1A protein, SEQ ID NO:37.
XX
DE Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KM non-typeable Haemophilus influenzae; NTH1; non-encapsulated;
KM recombinant production; Escherichia coli; antibacterial; vaccine;
KM human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KM detection; diagnosis.
XX
OS Haemophilus influenzae strain K1.
XX
FH Key Location/Qualifiers
FT Misc-difference 307
FT FT
XX
PN W0200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
XX
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
XX WPI; 2000-303789/26.
XX
DR N-PSDB; AAA52180.
XX
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 8; Fig 20A-R; 307pp; English.
XX

The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMW1A proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMW1A
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMW1A. The invention also discloses hmwA genes (AAA52175-A52198)

```



```

OY 179 VK-----RRSEYGYTYTINITIKLTD-----KGNIOIWLPOFKSDARVDLNRPTGG 227
      : : | : : ||| | ||| : : ||| :
Db 496 ISIAKGAIKFDIENTGSLNITTKSDSNHHITIKGNI-----TNRKGDNLITNNGDN 547
      : : | : : ||| | ||| : : ||| :
OY 228 T--YIGRSVDMCFYDGYSTNSS-----LEIRFODNPKSD-GKFYLRKINDTKEI 277
      : : | : : ||| | ||| : : ||| :
Db 548 TEIQIGN---ISOKEGMLTISSDKVNITERITIKAGVNGDSDSNEATSANLTIKTKE 604
      : : | : : ||| | ||| : : ||| :
OY 278 AVTSLLAG---KSLTPNGSLNIADASL-ETNNRITAVTMEISVPLCPGRLO 333
      : : | : : ||| | ||| : : ||| :
Db 605 KLTNDLNTSGFNKAETAKDNSNLFTIGDSDAGNTDARKVTFSNVKSISAS--DHNYT 662
      : : | : : ||| | ||| : : ||| :
OY 334 IDAKVE---NPEAGGYMGNIINVTFTSSOTL 361
      : : | : : ||| | ||| : : ||| :
Db 663 LNSKVETSGDITDTEGDNNTGLTITAKV 693
      : : | : : ||| | ||| : : ||| :

```

Search completed: December 4, 2002, 17:27:37
 Job time : 43 secs

This Page Blank (uspto)